APPLIED STATISTICS

A Journal of the
Royal Statistical Society

EDITOR
DONALD G. BEECH

ASSISTANT EDITOR
HAROLD S. BOOKER

VOLUME SEVEN

Oliver and Boyd Ltd.
London: 39A Welbeck Street, W.1
Edinburgh: Tweeddale Court

1958

CONTENTS (*continued*) *Page*

Notes and Comments 126

Meetings of Sections of the Royal Statistical Society 127

Book Reviews:

Statistical Sources for Market Research 130

Introduction to Statistical Reasoning, Philip J. McCarthy 130

Regression Analysis of Production Costs and Factory Operations Philip Lyle, revised by Leonard H. C. Tippett 132

Publications Received 132

No. 3. NOVEMBER

A Guide to Acceptance Sampling .. *Leonard H. C. Tippett* 133

Some Basic Principles of Sampling Inspection by Attributes *Hugo C. Hamaker* 149

The Lognormal Distribution in Quality Control *James Morrison* 160

Analysis of Variance of a Randomised Block Design with Missing Observations *William Alexander Glenn and Clyde Young Kramer* 173

Professor Hogben's 'Crisis'—A Survey of the Foundations of Statistics *D. V. Lindley* 186

Notes and Comments 199

Meetings of Sections of the Royal Statistical Society 200

Book Reviews:

Social Aspects of Prescribing, J. P. Martin 201

Statistical Methods in Quality Control, Dudley J. Cowden 202

An Introduction to Probability Theory and its Applications, Vol. I William Feller 203

Publications Received 204

STUDIES IN THE VARIABILITY OF TWO BLOOD-CLOTTING TESTS

PETER ARMITAGE and G. ILSLEY C. INGRAM*

Statistical Research Unit of the Medical Research Council,† and Department of Surgery, University of Edinburgh

In a study of the variability of the results of blood-clotting tests, interest centred on the coefficient of residual variation about regression on time. The coefficient was found to vary from one occasion to another; Dr Armitage and Dr Ingram here show how this variation was analysed with respect to some potentially relevant factors. They also describe the development of rapid significance tests for differences between means.

Introduction

When the clotting time of a sample of blood or plasma is measured it is customary to make a number of replicate readings and to calculate their mean. In clinical work, comparable readings are often obtained from a patient and a known normal subject. Since the precision of the mean clotting time depends on the variability of the individual readings, we thought it worth while to investigate this variability for two different tests in which replicate readings may be obtained one after the other: the thrombin clotting time of plasma (*thrombin time*) and Quick's one-stage prothrombin time test (*prothrombin time*).

The data which have been examined were obtained in the course of other work over a period of about two years. A 'retrospective' investigation restricts the information that can be obtained on the causes of variability, but has the advantage that the observers whose work has been analysed were not biased by knowing that the variability of their observations would be scrutinised.

We are principally concerned with the *variability* within groups of replicate readings, each group being made by one observer on a single sample of plasma. The purposes of the analysis were, first, to measure this variability and to study its relation to the mean clotting time; second, to see whether the variability was related to changes in technique or to clinical classification of the subjects; third, to compare the variabilities of these and other haematological tests; and, last, to construct a rapid test for the significance of differences between the means of two groups of replicate readings. The present data provide little information on the extent to which clotting times vary on the average between different observers making the test, between different subjects

* External Staff, Medical Research Council. Present address: Louis Jenner Laboratory, St. Thomas's Hospital, London, S.E.1

† London School of Hygiene and Tropical Medicine, Keppel St., Gower St., W.C.1.

from whom plasma is tested (although this is briefly considered) or between different times at which plasma obtained from a given subject is tested.

Methods

The tests (Biggs and Macfarlane[3]) were made at 37° C. on human venous plasma. To prevent spontaneous coagulation, the blood had been decalcified by treatment either with sodium citrate solution or with amberlite. In making the tests, the following modifications were introduced. First, the samples were also tested after the fibrinogen had been removed by clotting with 2–5 u/ml.* of purified human thrombin ('defibrination') and replaced by purified pooled human fibrinogen. Second, the tests (often with both whole and defibrinated plasma) were made with several dilutions of the brain reagent or the thrombin solution, which affect the mean clotting time and so give patient-control comparisons over a wide range of clotting times. Third, some of the samples were tested on the day on which they were obtained, whereas other samples were stored for varying periods at −20° C. before being thawed out and tested. For replicate readings, materials for each individual test were withdrawn from the stock solutions and placed in a warmed (37° C.) test tube immediately before use: materials for series of tests were not put out beforehand at 37° C. (There was some departure from this rule in the early tests, which may have affected the variability. Unfortunately it was not possible to investigate this because of differences in detail in the conditions of different tests, but at least the variability in these data was not obviously higher than in the remainder.) While being tested, some samples were held at room temperature and others in an ice-bath.

The first and second series of tests analysed below were all performed by one experienced observer, whereas the material of the third series was obtained from tests made by three different observers beginning work in this field.

Analysis of Thrombin Time Results (First Series)

The investigation was done on three series of test results, obtained over different periods of time. We shall describe in some detail the analysis of the thrombin time results from the first series, and then briefly report the conclusions drawn from the second series of thrombin times (this test was not used in the third series), and the three series of prothrombin times.

The first series consisted of 76 groups of replicate readings with between two and six readings per group, made on 23 samples of plasma from patients or from normal subjects. The mean thrombin time per group varied from 6 to 30 sec. A preliminary examination of the data suggested a tendency for the times recorded towards the end of a group

* Units of thrombin per millilitre of plasma (final concentration).

to be rather longer than those made at the beginning, presumably owing to changes in the reagents or in the reactions of the observer. If this were a real effect it seemed most useful to measure the residual variation after elimination of the trend: first because the total variation would presumably tend to be greater with a large than with a small number of readings, and secondly because the variation about the trend would be more relevant to the construction of a significance test for the comparison of two means, as the trend itself can be eliminated by experimental design (see below).

To test fairly quickly the significance of this regression on time, the readings in each group were ranked in order of thrombin time, and also according to order of observation. Rank correlation methods[7] showed a slight preponderance of high positive associations between these rankings and a deficit of high negative associations, especially in the groups of four, five, and six readings. The tendency was hardly significant, but as it accorded with previous experience and appeared in a more marked form in the prothrombin time results, it seemed safe to assume the effect to be real.

Accordingly, for each group of readings, we estimated the standard deviation about a linear regression on time. For this purpose, the readings were taken to be equally spaced in time; this was not strictly true for all the groups, but was considered an adequate assumption. For groups of two readings, of course, no estimate of variation about regression could be obtained and they were therefore not used. The data were thus reduced to 41 groups of replicate readings, on 19 different plasma samples from 18 subjects.

Denoting the thrombin times by x, and the (supposedly equally spaced) times of observation by t, the standard deviation of x about regression on t is estimated by

$$s = \theta \sqrt{\left\{ \frac{\mathrm{S}(x - \bar{x})^2 - \{\mathrm{S}x(t - \bar{t})\}^2 / \mathrm{S}(t - \bar{t})^2}{k - 2} \right\}}$$

where k is the number of readings in the group ($k = 3$, 4, 5 or 6), θ is a bias-correcting factor shown in Table I, and S denotes summation over the k readings. An explanation of this formula is given in the Appendix.

TABLE I

Factors used in the estimation of residual standard deviation about regression on time

Number of Replicates k	Bias Correction θ	Weight w
3	1·25	1·8
4	1·13	3·7
5	1·09	5·6
6	1·06	7·6

A general tendency for the residual standard deviation, s, to increase with the mean, $\bar{x}$, was immediately apparent. For each group we calculated the *coefficient of residual variation*,

$$z = 100s/\bar{x}\ \%$$

Fig. 1 shows a scatter diagram of z against $\bar{x}$. Clearly, z is virtually independent of $\bar{x}$, and a regression analysis confirms this visual impression. In this analysis (shown in Table II) each of the 41 values of z is weighted by a factor w depending on the number of readings, and given in Table I (see Appendix). The values of z and w are shown in

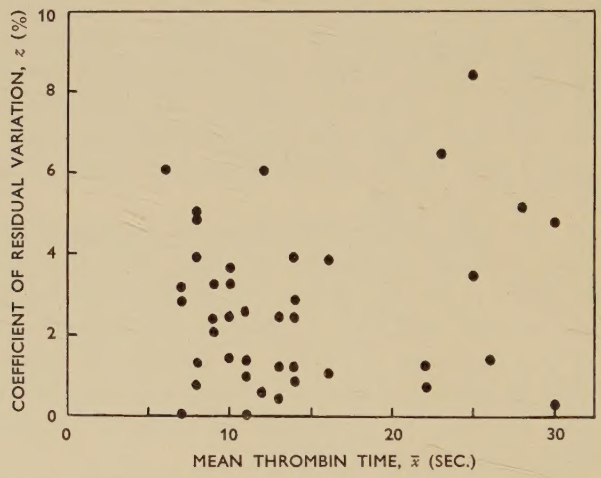


FIG. 1. Scatter diagram showing coefficient of residual variation and mean, for thrombin clotting time of plasma (Series 1).

TABLE II

Analysis of variance for the regression of coefficient of residual variation, z, on mean, $\bar{x}$, for thrombin time (Series 1)

	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratio
Due to regression ..	4·28	1	4·28	0·38
About regression ..	443·29	39	11·37	
	447·57	40		

columns (2) and (3) of Table IIIa. The mean square due to regression in Table II is less than that about regression, and is clearly non-significant. Now, the mean value of z is $\bar{z} = 2{\cdot}89\%$. The system of weighting is such that, if all the scatter in z were due purely to sampling variations (i.e. if there were no real heterogeneity in z), the expectation of the mean square about regression in Table II should be fairly close to $\bar{z}^2 = 8{\cdot}37$. In fact the mean square is 11·37, greater than $\bar{z}^2$ but not

significantly so. There may be some unexplained heterogeneity in the coefficient of residual variation, but the above analysis provides no strong evidence for it.

We considered, however, that certain explicit factors might be associated with the real fluctuations in z, and decided to examine these in more detail. The possible causes of heterogeneity in z, which could be investigated with the present data, seemed to be

(*a*) differences between individual plasma samples, in particular those from subjects in different clinical groups;

(*b*) whether or not the sample had been defibrinated; and

(*c*) length of storage at −20° C. before test.

The data are shown in Table IIIa. Information on the effect of (*b*) could be obtained *within* plasma samples, since groups of readings were

TABLE IIIa

Data obtained from first series of thrombin times

Clinical Group	Sample Number	(1) Storage at −20° C. (days)	(2) Whole Plasma		(3) Defibrinated Plasma		(4) Difference (W − D)		(5) Pooled Value	
			w_W	z_W	w_D	z_D	w_d	d	w	z
(i) Thrombotics	1	15	1·8	6·0	—	—	—	—	1·8	6·0
	2	1	5·6	2·0	5·6	2·5	2·8	−0·5	11·2	2·2
	3	2	7·6	6·0	3·7	8·3	2·5	−2·3	11·3	6·8
	4	9	—	—	3·7	1·2	—	—	3·7	1·2
	5	5	—	—	3·7	3·9	—	—	3·7	3·9
	6	—	3·7, 3·7 } 7·4	5·0, 2·4 } 3·7	—	—	—	—	(7·4	3·7)
(ii) Bleeders	7	1	5·6	4·8	5·6	1·4	2·8	3·4	11·2	3·1
	8	8	1·8	1·2	1·8	6·4	0·9	−5·2	3·6	3·8
	9	3	—	—	3·7	2·8	—	—	3·7	2·8
	10	5	1·8, 1·8, 1·8 } 5·4	0·7, 0·9, 3·8 } 1·8	1·8	0·3	1·4	1·5	7·2	1·4
	11	4	1·8, 1·8, 1·8 } 5·4	3·9, 0·0, 1·0 } 1·6	1·8	4·7	1·4	−3·1	7·2	2·4
(iii) Liver disease	12	16	1·8	0·4	1·8	0·7	0·9	−0·3	3·6	0·6
(iv) Normal controls	13	1	3·7	3·2	5·6	1·3	2·2	1·9	9·3	2·1
	14	2	7·6	2·8	3·7	3·4	2·5	−0·6	11·3	3·0
	15	2	1·8, 1·8, 1·8 } 5·4	0·0, 3·6, 0·8 } 1·5	1·8	1·4	1·4	0·1	7·2	1·4
	16	4	1·8, 1·8, 1·8 } 5·4	3·1, 2·3, 2·4 } 2·6	1·8	5·1	1·4	−2·5	7·2	3·2
	17	5	—	—	3·7	2·4	—	—	3·7	2·4
	18	16	1·8	0·5	1·8	1·2	0·9	−0·7	3·6	0·8
	19	—	3·7, 3·7 } 7·4	1·3, 3·2 } 2·2	—	—	—	—	(7·4	2·2)

obtained on the same samples with both whole and defibrinated plasma. The effect of (*c*) had to be assessed from comparisons *between* samples, since the different readings on each sample of plasma invariably corresponded to the same, or almost the same, length of storage time. The following clinical classification of plasmas was adopted (groups of two readings discarded): (i) thrombotics (six samples from different patients, only five of which are included in the between-sample analysis in Table IIIb, since no storage time was recorded for the sixth); (ii) bleeders (five samples from different patients); (iii) liver disease (one sample); (iv) normal controls (seven samples from six subjects, six of which, from different subjects, are included in between-sample contrasts in Table IIIb).

TABLE IIIb

Analysis of variance of the coefficient of residual variation, z, of thrombin time (*Series* 1)

	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratios
Between plasma samples:				
Between clinical groups, corrected for effect of storage time	70·70	3	23·57	1·41
Regression on storage time ..	18·84	1	18·84	1·13
Residual, within groups ..	199·60	12	16·63	1·99* (1·00)
Within samples:				
Defibrination	1·78	1	1·78	0·21
Samples × defibrination ..	103·58	11	9·42	1·12
Residual	58·34	10	5·83	0·70
	Theoretical variance		8·37	(1·00)

* $0{\cdot}01 < P < 0{\cdot}05$.

The analysis of variance of z is shown in Table IIIb. It may be useful to indicate briefly the derivation of the various items.

(i) The within-samples residual mean square is obtained from the six samples of whole plasma for which two or more values of z were available [see column (2) of Table IIIa]. For each of these samples the weighted sum of squares, $\Sigma w(z - \bar{z})^2 = \Sigma wz^2 - (\Sigma wz)^2/\Sigma w$, is calculated, and the sum of these contributions gives the entry 58·34 in Table IIIb. Since there are 16 values of z from 6 samples, this sum of squares has 10 degrees of freedom (d.f.). The residual mean square, 5·83, is not significantly high; it is, in fact, rather lower than its expected value, 8·37. There is, therefore, no evidence of heterogeneity in z *within* plasma samples.

(ii) There are 12 samples for which values of z are available with both whole and defibrinated plasma. These values, given in columns (3) and (4) of Table IIIa, are denoted by z_W and z_D [using weighted means of the repeated values referred to in (i)], and the corresponding weights by w_W and w_D (using sums of the weights of the repeated

values). We then obtain the difference, $d = z_W - z_D$, and its weight, $w_d = w_W w_D/(w_W + w_D)$, which are given in column (4) of Table IIIa. The sum of squares (on 1 d.f.) for the effect of defibrination is $(\Sigma w_d d)^2/\Sigma w_d$, the summation being over the 12 available samples, and the sum of squares corresponding to the samples × defibrination interaction (on 11 d.f.) is $\Sigma w_d d^2 - (\Sigma w_d d)^2/\Sigma w_d$. Neither of the two mean squares is significant, and we conclude that there is no evidence that defibrination has any overall effect on z, or that its effect varies from sample to sample.

(iii) In view of the non-significance of the defibrination main effect and interaction, in this and in all similar analyses, we thought it justifiable to ignore the distinction between whole and defibrinated plasma in the remainder of the analysis. Had this not been possible the analysis would have been non-orthogonal and much more complex. For each sample, then, the pooled values of z and w given in column (5) of Table IIIa are obtained as follows:

$$z = (w_W z_W + w_D z_D)/(w_W + w_D) \quad \text{and} \quad w = w_W + w_D$$

For 17 of the 19 samples, values of the storage time are available, and are given in column (1) of Table IIIa. The other two samples were omitted from the between-samples analysis, which consisted of an analysis of covariance of z between the four clinical groups, corrected for concomitant variation in the storage time, and using the weights w. The sum of squares on 1 d.f. for regression of z on storage time is calculated *within* clinical groups, and is non-orthogonal to that on 3 d.f. for the corrected differences between clinical groups. These two sums of squares should not, therefore, be added.

From Table IIIb we see that there *is* significant heterogeneity in z from sample to sample ($P < 0{\cdot}05$). In view of this the other mean squares in the analysis must be tested against the residual mean square *between* samples. The variation in z between samples is not clearly related either to clinical group or to storage time.

To summarise the analysis presented in Tables IIIa and IIIb, then, we have found evidence of real variation in z between plasma samples, but no evidence that any of the tested factors has any effect on z.

Analysis of Thrombin Time Results (Second Series)

This series was examined for evidence of the possible effect on z of

(*a*) the method of decalcification (citrated or amberlite), and

(*b*) the temperature of the plasma at the bench (ice-bath or at room temperature).

There was no evidence of any heterogeneity in z, either between plasma samples, or due to the factors (*a*) and (*b*) mentioned above. As in Series 1, z appeared to be independent of $\bar{x}$, and the mean value of z was $\bar{z} = 3{\cdot}06\%$, a value close to that obtained in Series 1.

Analysis of Prothrombin Time Results (First and Second Series)

These observations were made over the same period, and mainly on the same subjects, as the two corresponding series of thrombin time results. Fig. 2 shows a scatter diagram of z against $\bar{x}$ for the Series 1

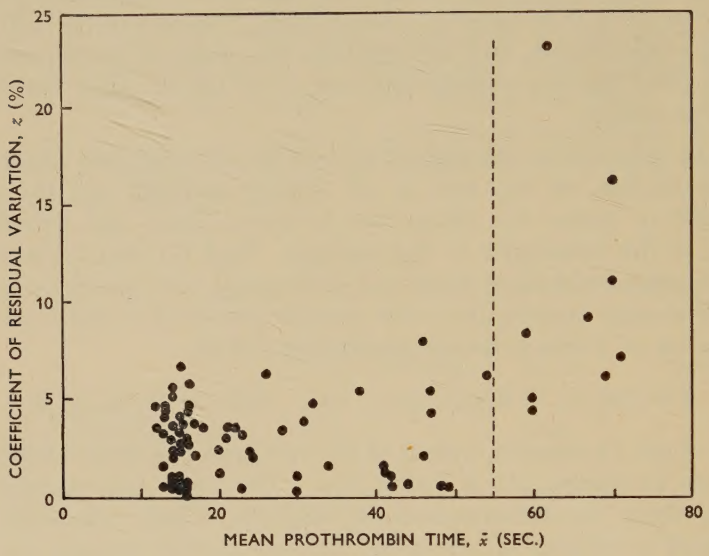


FIG. 2. Scatter diagram showing coefficient of residual variation and mean, for one-stage prothrombin time (Series 2).

results; after the omission of groups of two readings there were 75 groups of three to six replicate readings, from 25 plasma samples. The apparent tendency for z to increase with $\bar{x}$ was confirmed by a regression analysis, but Fig. 2 suggests that the increase in z occurs mainly, or solely, at high values of $\bar{x}$. Such high prothrombin times, obtained with very dilute brain reagent, can be avoided in practice, and it seemed relevant to examine the values of z for those groups of readings with a mean time not greater than 55 sec. (i.e. to the left of the vertical line in Fig. 2). A regression analysis on the remaining 66 groups of replicate readings reveals no significant association of z with $\bar{x}$. The precise value chosen as an upper limit for $\bar{x}$ is of course arbitrary, and a considerably greater number of observations would be needed before the relationship of z to $\bar{x}$ could be determined at all accurately.

The analysis proceeded along the same lines as that for thrombin times. The only important effect was a difference between clinical groups, which reached the 1% level of significance. The principal difference here was between the thrombotics and other groups (see Table IV). This was not confirmed in the analysis of the second series, and we personally believe it to be a spurious result, unrelated to the diagnosis of the patients. An occasional difference in variability may be due to the inadequacy of the assumption of a linear trend of clotting time with the time at which the test is performed. Any departure from

a linear trend, or irregularity in the intervals at which replicate tests are made, will tend to increase the estimated residual variation, and this effect may well have been more marked on some occasions than on others.

With the second series of prothrombin times, as in the corresponding thrombin time analysis, it was possible to compare the effects on z of decalcification and standing temperature at the bench. The effects of these factors were again non-significant, but the residual mean square was significantly higher than its expected value. Again, there was evidently some unexplained heterogeneity in z.

Analysis of Prothrombin Time Results (Third Series)

In the course of other work, prothrombin time tests have been performed daily on batches of plasma samples from a number of post-operative patients and on control subjects, and occasionally on pooled normal plasma stored at −20° C., usually with four replicate readings per plasma sample. Tests were done with both dilute and concentrated brain reagent (dilution of the brain reagent lengthens the prothrombin time). An analysis of variance, for each day of testing and for each brain concentration, enabled the differences in prothrombin time between different plasmas, and those due to general deterioration throughout the experiment, to be eliminated. The square root of the residual mean square in such an analysis should be comparable with the estimated standard deviation about regression in the studies reported above. For each of 57 days, and for each brain concentration, an estimated coefficient of residual variation, z', was obtained by dividing the root mean square by the mean prothrombin time.

No association between the coefficient of residual variation, z', and the mean prothrombin time, $\bar{x}$, was detectable for either brain concentration. On 51 of the 57 days, however, z' was greater for concentrated than for dilute brain over a range of clotting times of 12–56 sec., which is similar to the restricted range of 12–54 sec. used for the analysis of the first series. Since dilution of the brain gives a longer prothrombin time, this finding seems to contradict the positive association between z and $\bar{x}$ reported for the first series, over the wider range of 12–140 sec. We do not know why this should be. Perhaps a larger series than that shown in Fig. 2 would reveal a tendency for z to decrease slightly as $\bar{x}$ increased to, say, 50–60 sec., the rise in z taking place only at higher values of $\bar{x}$.

With concentrated brain (giving a mean prothrombin time of 18·1 sec. over different batches of this reagent), the values of z' ranged from 1·5% to 5·7%, with a mean of 3·16% and a standard deviation of 0·81%. For dilute brain (mean prothrombin time, 38·4 sec. over different batches), z' ranged from 1·2% to 6·7%, with a mean of 2·48% and a standard deviation of 0·95%. The variation in z' from day to day is considerably greater than would occur by chance, for each value of z' is based on between 20 and 30 degrees of freedom.

The standard error of values of z' based on, say, 25 degrees of freedom, would be estimated as $\bar{z}'/\sqrt{(2 \times 25)}$, i.e. $0{\cdot}1414\bar{z}'$, or 0·45% for concentrated brain and 0·35% for dilute. The standard deviations of z' are therefore two or three times their expected values.

On each day on which tests were performed on patients, two normal subjects (ostensibly healthy male medical students) were also tested. We were thus able to measure the variation in mean clotting time between plasmas from different normal subjects, for the same observer and on the same occasion. The data from 56 differing pairs of normal subjects shows that, for both brain concentrations, the standard deviation of the difference between mean prothrombin times for the paired subjects (with four replicate readings per subject) is about 9% of the combined mean time. The coefficient of residual variation discussed in this paper may be taken to be about 3%, and its contribution to the standard error of the difference between two means, each of n replicate readings, is therefore $3\sqrt{(2/n)}\%$ of the mean time. For $n = 3$, 4, and 5, this contribution amounts to 2·4%, 2·1%, and 1·9%, respectively. These values are considerably lower than the total standard error of 9%, which shows that the component representing real difference in mean time between subjects is much more important than the residual variation. The total value, of about 9%, will therefore be hardly affected by changes in the number of replicate readings per subject.

Comparison of Two Groups of Readings

In blood-clotting investigations, comparisons are often made between the clotting times of two different samples, which may be samples of the same plasma treated by two different methods, or samples from different plasmas. Replicate clotting tests may be performed for each sample, and it is helpful to be able to decide when the difference between the means may readily be attributed to chance fluctuations.

Such a comparative experiment can be designed so as to eliminate any general trend in clotting times due to deterioration of reagents or observer facilitation or fatigue. The appropriate designs are tabulated by Cox.[5] For instance, if the two samples are denoted by 1 and 2, and four replicate readings are to be made on each sample, the tests may be done in the order

2 1 1 2 1 2 2 1

The allocation of the numbers 1 and 2 to the two plasmas should be made at random. These designs have the property that the comparison of the two means is unaffected by a linear or quadratic trend (or, for some designs, a cubic). For odd numbers of replicates no design exists which completely eliminates even a linear trend. However, since the trend is not normally very pronounced it would be adequate to adopt a design which is nearly balanced. These also are given by Cox.

The theoretical value of these designs depends of course upon the equal spacing of the observations in the time sequence. In practice it is sufficient to avoid irregular pauses between tests, and to ignore the variations necessarily produced by differences in the clotting times of the tests themselves. It is also wise to work with fairly small groups of samples so that the interval between replicate readings in a given sample will be small. Non-linearity in the temporal regression will then be unlikely to invalidate the suggested significance tests.

In the analysis of such an experiment the relevant measure of random error would be the variation about the regression on the order of testing. This could be estimated separately from each experiment, but for an approximate test it would be convenient to use an estimate of the coefficient of residual variation derived from the studies reported above. The various estimates, with their standard errors, are shown in Table IV. The standard errors were calculated as $\bar{z}/(\Sigma w)^{\frac{1}{2}}$, where

TABLE IV

Different estimates of the mean coefficient of residual variation of thrombin time and prothrombin time

	Mean value $\bar{z}$	Standard Error of $\bar{z}$	Comments
THROMBIN TIME:			
Series 1	2·89	0·37	Heterogeneity between samples
Series 2	3·06	0·26	
PROTHROMBIN TIME:			
Series 1			
Group (i) (Thrombotics) ..	3·90	0·48	
Groups (ii)–(iv) (Others) ..	2·44	0·21	
Series 2	2·75	0·24	Heterogeneity within samples
Series 3			
Concentrated brain ..	3·16	*	
Dilute brain	2·48	*	

* Standard errors omitted because values of z did not fluctuate randomly (there was a fluctuating time trend but insufficient data to obtain an adequate estimate).

no heterogeneity in z was detected, and otherwise as $(m.\,sq./\Sigma w)^{\frac{1}{2}}$, where $m.\,sq.$ denotes a mean square of z (either within or between samples) which significantly exceeded the theoretical variance.

We have taken, from Table IV, the weighted mean values of $\bar{z}$: 3·0% for thrombin time, and 2·8% for prothrombin time. The standard error of the difference between the means of two samples of n_1 and n_2 observations is then estimated as

$$(0{\cdot}030)\bar{x}\sqrt{(1/n_1 + 1/n_2)} \text{ for thrombin time}$$

and $$(0{\cdot}028)\bar{x}\sqrt{(1/n_1 + 1/n_2)} \text{ for prothrombin time}$$

where $\bar{x}$ is the mean of all the clotting times obtained from both

samples. The difference between two means would be significant at about the 5% level if it exceeded twice its standard error. These formulae for the standard errors would, of course, be valid only for data whose variability was comparable with that of the original data analysed above.

A significant difference between the means of replicate readings for a patient's plasma and a control plasma will not necessarily indicate abnormality in the patient's plasma, since variation in clotting time is observed between different normal subjects. The relevant question is likely to be whether the difference in means is greater than would usually occur between means of samples from two normal subjects. From the analysis of the third series, the standard error of the difference in means (at least for prothrombin time) may be taken as approximately $0{\cdot}09\bar{x}$ (the effect of varying numbers of replicates being negligible as a first approximation).

If the test plasma is expected to differ from the control in one direction only (as in deficiencies of factor V or factor VII, when the patient's clotting time will be expected to be longer than the control), the difference between the mean prothrombin times of the two plasmas reaches significance at the 5% level if it bears a ratio to its standard error of more than 1·64, and at the 1% level if the ratio exceeds 2·33. In the investigation of clotting deficiencies, Macfarlane and Biggs[8] (p. 12) recommend that the brain suspension used for the one-stage prothrombin time should clot normal plasma in 12–20 seconds, and say that 'the patient's clotting time is abnormal if it exceeds that of the normal by more than 2 or 3 seconds'. Our results indicate that the clotting time of one normal plasma might exceed that of another by as much as 1·9 sec. when the mean of their clotting times was 12 sec., and by as much as 3·2 sec. when the mean was 20 sec., at the 5% significance level. These figures agree very closely with the recommendations of Macfarlane and Biggs.

If the test plasma prothrombin time might be either longer or shorter than the control plasma, an observed difference will be significant at the 5% level if it exceeds 1·96 times the standard error (i.e. $0{\cdot}18\bar{x}$), and at the 1% level if it exceeds 2·58 times the standard error (i.e. $0{\cdot}23\bar{x}$).

Discussion

The general conclusion from these studies is that, after elimination of a slight tendency for readings to increase with time, both the clotting tests show a coefficient of variation of about 3%. We have found no clear indication of any technical features of the tests affecting this figure. Similar values have been reported by other authors,[1,10] and by ourselves in a previous paper.[6] On the present evidence, the random errors in these two tests compare favourably with those of whole blood clotting tests,[4,6] and indeed with those of other current haematological techniques.[4]

The simple significance tests for the difference between two means have been useful in our own routine work. The test based on differences between means from two normal subjects has been used for comparisons of samples from different plasmas, particularly in comparing the blood of a patient suspected of some blood disorder with that of a normal control. The test based on residual variation is useful for comparing the effects of different methods of treating aliquots from one sample of plasma. In particular, the residual variation is relevant in various assays, such as those of heparin, factor V, and factor VII, which may use prothrombin time as a response.

Acknowledgements

We should like to thank Miss Irene Allen for computing assistance and Mrs M. G. Young for preparing the diagrams.

APPENDIX:

ESTIMATION OF COEFFICIENT OF RESIDUAL VARIATION

If k replicate clotting times, x, follow a linear regression on the time of observation, t, the variance of x about the regression line is estimated by

$$s_0^2 = \frac{\mathrm{S}(x - \bar{x})^2 - \{\mathrm{S}x(t - \bar{t})\}^2/\mathrm{S}(t - \bar{t})^2}{k - 2}$$

where S denotes summation over all the readings $x_1, x_2, \ldots, x_k$ at times $t_1, t_2, \ldots, t_k$.

The statistic s_0^2 is an unbiased estimate of the residual variance, σ_0^2; that is, in repeated sampling it will on the average be equal to σ_0^2. The statistic s_0 is a biased estimate of the residual standard deviation, σ_0, being on the average rather too low. An unbiased estimate is provided by

$$s = \theta s_0,$$

where θ is the factor given in Table I. The sampling variance of s is σ_0^2/w where w is given in Table I. These factors may be obtained from Table 35 of Pearson and Hartley.[9] Reading along the line corresponding to $\nu = k - 2$, θ is the reciprocal of the entry in column (2) of this Table, and w is the square of the ratio of the entries in columns (2) and (3).

We have used the estimates of the coefficient of residual variation, z $(= 100s/\bar{x})$. The sampling error of z is almost entirely determined by that of s, since the error in $\bar{x}$ is relatively small. If this latter component is ignored the sampling variance of z is ζ^2/w, where ζ is the expected value of z. In an analysis of variance of different values of z, each value should be weighted in inverse proportion to its sampling variance, i.e. the w may be taken as weights. On the null hypothesis that the values of z differ by sampling fluctuations from a common value, ζ, the mean squares in the analysis should be estimates of ζ^2, which may be estimated by the square of the mean value of z.

The tests of significance used in the analyses of variance should be regarded as approximate. In the first place, the sampling error of $\bar{x}$ is ignored, as in the argument above; secondly, we have ignored the error involved in replacing the theoretical variance, ζ, by its estimated value, $\bar{z}$; thirdly, the validity of the tests will be slightly affected by the non-normality of the distributions of z; fourthly, the factors θ and w are strictly appropriate only when x is normally distributed about the regression on t. The effects of these approximations should not be serious. The more customary method of handling data of this type[2] would be to work with the logarithm of s or z. This approach was not adopted here, because the necessary rounding-off of clotting times meant that occasionally s (and therefore z) was zero, and a logarithmic transformation would have caused difficulties. A modified transform, $\log(s + \text{constant})$, could have been used, but the choice of the constant would have been rather arbitrary.

An alternative approach would have been to work throughout with $\log x$. We might then have expected the standard deviation of $\log x$, about regression on t, to remain constant, apart from sampling errors. This approach was not adopted because the values of s^2 had been calculated before the general pattern of the results was apparent, and there seemed little point in extensive recalculation. The standard deviation of $\log x$ might have been more nearly normally distributed than was z, and the analyses of variance consequently more reliable. It is difficult to determine whether this would be so, without more information about the form of the distribution of x than is available from small samples.

REFERENCES

[1] Aggeler, P. M., Howard, J., Lucia, S. P., Clark, W. and Astaff, A. (1946). 'Standardization of the Quick prothrombin test', *Blood*, **1**, 220.

[2] Bartlett, M. S. and Kendall, D. G. (1946). 'The statistical analysis of variance-heterogeneity and the logarithmic transformation', *Suppl. J. R. Statist. Soc.*, **8**, 128.

[3] Biggs, R. and Macfarlane, R. G. (1957). *Human Blood Coagulation and its Disorders*, 2nd edition, pp. 402, 416. Blackwell, Oxford.

[4] Biggs, R. and Macmillan, R. L. (1948). 'The errors of some haematological methods as they are used in a routine laboratory', *J. Clin. Path.*, **1**, 269.

[5] Cox, D. R. (1951). 'Some systematic experimental designs', *Biometrika*, **38**, 312.

[6] Ingram, G. I. C. and Armitage, P. (1952). 'The effect of penicillin and of streptomycin on blood coagulation in normal subjects', *J. Pharm. Pharmacol.*, **4**, 1047.

[7] Kendall, M. G. (1955). *Rank Correlation Methods*, 2nd edition. Griffin, London.

[8] Macfarlane, R. G. and Biggs, R. (1955). *The Diagnosis and Treatment of Haemophilia and its Related Conditions*. Med. Res. Coun. Lond. Memo. No. 32; HMSO, London.

[9] Pearson, E. S. and Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, Vol. I. Cambridge University Press.

[10] Rhees, M. C., Ellerbrook, L. D. and Brown, D. V. (1955). 'Plasma prothrombin time. A method having a photoelectric end-point with controlled temperature', *Amer. J. Clin. Path.*, **25**, 453.

OFFICIAL STATISTICS OF ECONOMIC TRENDS

RITA J. MAURICE
University College, London

In the last few years a number of important changes have been made in the government's programmes for the collection of economic statistics. In this article these changes are summarised in the light of the need to detect changes in the level of economic activity.

The 1944 *White Paper on Employment Policy*[1] recognised that no full employment policy could be successful unless based on sufficient information on economic trends. It therefore listed the principal classes of statistics, in addition to those available before the war, which had to be collected. The 1947 Statistics of Trade Act gave the government wide powers for the collection of compulsory information, and government statistical services were much expanded. However, in his budget speech of April 1956 the Chancellor of the Exchequer (Mr Macmillan) commented on the difficulties of 'looking up a train in last year's *Bradshaw*'[2] and the necessity of improving the statistics both 'in form and timing'. On 1st August 1956[3] he outlined the government's future programme and the improvements intended to fill this need.

Some of the post-war changes are here examined in the light of two American studies of short-term forecasting. The first is a statistical analysis examining the regularity with which fluctuations in some hundreds of economic series were related to the general level of economic activity in the United States before 1948.[4] This study resulted in a tentative list of twenty-one reliable indicators but these were inevitably limited to the data available for analysis. Some other indicators of activity were discussed at a conference on short-term economic forecasting held in 1951.[5] In November 1955 the *Economic Bulletin for Europe* discussed the availability of certain indicators in Western Europe,[6] and reference is also made to these series. The main subject fields of economic statistics are discussed in turn.

Manpower

Before the war administration of unemployment insurance provided monthly figures of unemployment and an annual figure of employment. With the introduction of the National Insurance Scheme in July 1948 the coverage of these series was extended to the entire working population. After 1939 these series were supplemented by compulsory direct returns by employers (L returns) of the employment they provided. Since 1945 these returns have continued on a monthly basis, as required in the *White Paper*, from all employers with more than ten workers

in manufacturing industries, but in 1948 and in October 1956 smaller employers were relieved of the obligation. Returns are made also by a sample of employers in distribution and in some service trades, but coverage is not complete and the sample may be biased, as it was originally obtained from those who had recorded at Employment Exchanges their need for workers. The burden of these returns has also been lightened since November 1956 by reductions in the size of the sample. Information is available from other sources on employment in agriculture, building, and nationalised industries. The *White Paper* also required information on prospective employment, which was collected quarterly on the 'L returns' but never published. This enquiry has gradually disappeared from the returns and the last question on the subject was asked in August 1956.

Thus the information required by the *White Paper* has been collected, but in practice it appears to be insufficient. Turner[7] found that in 1952 unemployment registration failed to reflect 'involuntary idleness' fully because it recorded only part of those *partially* unemployed, and workers failed to register if their prospects of re-employment were slight (e.g. part-time workers, married women) or if they were ineligible for unemployment benefit. Similarly, payroll employment of the 'L returns' was an 'inadequate index to loss of work because most of the workers affected remained at least nominally possessed of jobs'. Turner concluded that series of total hours worked and short-time and over-time are more significant of trends in employment. This conclusion is supported by Moore, who found that series of employment and unemployment were reliable indicators of activity but that they moved roughly coincidentally with the general level of activity whereas average hours worked per week moved in advance of it.

The Ministry of Labour has recently extended its information in this direction. The 'L returns' from manufacturing industry have since 1945 provided quarterly information on short-time and overtime working by operatives. Publication of this information was started in 1952, giving the number of hours and the number of operatives involved. In January 1957 a new series was started, consisting of monthly numbers of manufacturing operatives on short-time. Extensive short-time working prompted this information but it would seem less satisfactory in pointing to future declines in activity than the selection of overtime hours made by the American Bureau of Labour Statistics.[8] Information on hours worked is still limited to the voluntary six-monthly earnings enquiries of an uneven sample in manufacturing, building, road transport, and public utilities. Figures are published of average hours worked by wage earners, but are subject to a five-month delay; the key figures are available in four months.

Another leading indicator in this field is the number of job vacancies registered.[6] These are not published in this country, although an approximation may be obtained from figures of placings and unfilled vacancies. Except for limited periods when all vacancies had to be

filled through employment exchanges, these figures do not have complete coverage, although movements in them give useful indications of trends in the demand for labour.

Production

The 1944 *White Paper* required the institution of an annual census of production, which was started in 1948. It has, however, been found unnecessary to take a census in full detail each year and censuses relating to 1949 and 1950 were much simplified. In 1952 and 1953 the information collected was further reduced by requiring returns from only a sample of firms, and the Verdon Smith Committee in 1954 recommended that the lower exemption limit for a full census should be raised and that a complete census should be taken only every third year. The Committee suggested that the interval might be extended to five years after 1960 and in the event four years are to elapse before the next full census in 1958.

Between censuses monthly figures of production, and if possible orders on hand, were to be obtained from large firms, trade associations, and public institutions. In addition to the excise data previously available, much information was available on production during and after the war from the operation of controls. In recent years, however, controls have been abandoned and special enquiries have become necessary to maintain the series. The Political and Economic Planning report[10] described the position at the end of 1956 as follows:

> 'Owing to the way the statistics originated, there are a number of gaps in the monthly information. Only quarterly output figures are at present available for some engineering industries, and for some household durables: for example, carpets, furniture, pottery, and hosiery, and boots and shoes. There are at present no up-to-date output statistics for a few industries such as scientific instruments, heavy electrical equipment and household textiles. Only a poor sample enquiry exists for clothing.'

Mr Macmillan stated that the programme was to be expanded to obtain monthly figures from all of the most important industries and more frequent and new returns from a number of others. The first of the new monthly enquiries, for furniture output, began in October 1956. Monthly enquiries for pottery, carpets, and wood chipboard followed in January 1957, hosiery in February, scientific instruments in April, and footwear in July.

In 1948 an interim monthly index of industrial production was started, incorporating some 400 series. When the 1948 census information made a more accurate index possible, the present index using 1 300 series was introduced in 1952. In July 1956 the interpretation of this series was made much easier by the publication of a seasonally adjusted index.

Turner found production figures to be a more reliable indicator in the 1952 cotton recession than either employment or unemployment and Moore included the production index in his chosen list. However, this series did not 'lead' as did the allied series for new orders.

Although the *White Paper* required, if possible, monthly figures of orders, at the time of the Chancellor's statement official monthly figures of orders were limited to a series on metal-working machine tools begun in March 1955. Quarterly figures on domestic furniture were begun in January 1953, and became monthly in December 1956. Such series are difficult to collect and interpret because of duplication, possible cancellation, and seasonal variation. In addition, some industries carry sufficient stocks to enable immediate delivery. However, the ECE report considered the major obstacle to their use to be delay in publication (three and two months respectively for the series mentioned above). In the United States aggregate figures of new orders and the stock of orders are available after six weeks. Mr Macmillan promised an extension of figures on orders as well as on production and there is clearly considerable scope for improvement here.

Stocks

Stocks also provide a sensitive indicator of economic change, and were required monthly by the 1944 *White Paper*. Figures are published monthly for a number of industrial materials (other than foodstuffs and fuel), being obtained at the same time as the production figures. In the aggregate, however, there are no indices corresponding to the monthly index of production.

Manufacturer's stocks have been collected annually at the censuses of production, but these are subject to delay. At the end of 1953 a quarterly enquiry was started of a large number of manufacturing companies, but publication began only in January 1956. The sample is not a probability sample and the enquiry is voluntary, so that there is a possibility of bias even though the sample was strengthened at the beginning of 1956 by the inclusion of new companies. However, the larger sample has made possible a breakdown for broad industrial groups and according to the state of production. From the beginning of 1956 quarterly indices of materials held outside manufacturing industry have been compiled from the available commodity series. A new index of paper stocks was started in the middle of 1956.

For finished products the position has been very unsatisfactory since only about a quarter of the total stocks are held by manufacturers. The 1950 census of distribution and services provided information on stocks, but the next enquiry in this series is the sample enquiry for 1957. Between these enquiries the only information published on wholesale stocks was the monthly series for textiles and clothing started before the war. For retailers, a sample of department stores and multiple shops also provided figures before the war and continue to do so. This sample was strengthened in 1955 but no new stock indices were started.

Independent retailers have never made returns of stocks. Improvements in stock figures have to wait upon the completion of a satisfactory programme for retail sales series (see below).

Reporting problems again arise here, especially for small retailers, because totals of stocks may be compiled only infrequently. However, this may be overcome by deriving figures from purchases and sales or by using estimates. As with production and orders, seasonal variation makes stock figures difficult to interpret, and figures over several years are needed. But, in this case also, the obstacles are not insurmountable and in the United States aggregate stock data for manufacturing, wholesaling, and retailing are available within about six weeks of the end of the period.

Mr Macmillan recognised the gaps in the information on stocks and promised a compulsory annual sample enquiry of distribution and services, together with voluntary quarterly information. The first annual enquiry related to 1956 and a quarterly enquiry for wholesale distribution started for the last quarter of 1956. In addition all companies within the scope of the 1956 enquiry into distribution and services returned information on their kind of business in order to enable reliable estimates to be made of totals.

Fixed Investment

The 1944 *White Paper* envisaged regular information on capital expenditure projected by public authorities and private industry. Realised investment was asked for in censuses of production for 1948 and 1949, but quicker results were required and a voluntary sample enquiry was used to obtain information for 1949 and 1950. From 1951 onwards the census of production was again used but in July 1955 a quarterly enquiry was started of a sample of companies in manufacturing, distribution, transport, and services. However, the sample was proportionately smaller outside manufacturing, and fixed investment in distribution and services was not asked for at the 1950 census of distribution. The Chancellor noted the need for such information and at the end of 1956 the sample from distribution and services was increased. Also the new annual and quarterly enquiries mentioned above relate to fixed investment as well as to stocks if no return has already been made. The first annual figures also relate to 1956 but the quarterly investment figures for the wholesale distributive trade date from the end of 1955. This topic is also included in the 1957 census of distribution.

Both of the quarterly enquiries also ask for investment intentions, which were previously assessed from controls. As yet surveys of investment intentions have nowhere been very accurate but the report on short-term forecasting[5] found no serious bias in spite of large individual errors. Accuracy proved to be greater for larger firms, and Dow[11] concluded that such surveys had proved themselves useful in making investment forecasts, provided that they were not used mechanically.

A particular field of investment which yields useful 'forward-looking' series is that of building. Both building authorisations and starts are useful here. Industrial Development Certificates issued by the Board of Trade provide information since 1948, and since the end of building controls in 1954 these have been the basis for quarterly series on authorisations of industrial building for manufacturing industry. Although not all buildings authorised are built, a follow-up provides figures of starts and completions. Housing returns to local authorities have long provided information on dwellings started, under construction, and completed. A quarterly sample enquiry by the Ministry of Works started at the beginning of 1955 provides information on the value of all building work carried out.

Private and industrial construction contracts formed two of Moore's eight selected leading series and Bassie[5] recognised building contracts as one of the most dependable leading series. After the war there was published for the UK a quarterly series on new contracts for residential construction, but this was discontinued in November 1954 when building licences were no longer required. Mr Macmillan proposed in future to improve the building statistics with an analysis of figures on new contracts, and this should prove an important indicator. Questions were added to the Ministry of Works enquiry for the third quarter of 1956.

Foreign Trade

Moore did not include in his list of reliable indicators a series on external trade, but for this country it is obviously a field of the first importance. Adverse movements must be detected and acted upon quickly in order to safeguard our small reserves.

Monthly figures of the value of exports and imports passing through the ports, derived quarterly series of volume, and monthly price and terms of trade series were all available before 1944. In addition, the *White Paper* required quarterly estimates of foreign capital movements and the balance of foreign payments, but Mr Macmillan did not commit himself to as much detail as this.

Estimates of the balance of payments were made explicitly in the 1947 *National Income White Paper* and since February 1948 half-yearly *White Papers* have been published on the balance of payments. These have provided an increasing number of detailed breakdowns of the figures, and the reliability of most of the main categories is described by the Central Statistical Office[12] as 'fair'. However records for some 'invisible' items are described as 'inferior' and the inadequacy of data on long- and short-term capital movements is a large factor in the need for a 'balancing item' which in 1955 was initially as high as £150 million. (In 1955 the gold and dollar reserves dropped by £229 million.)

The Chancellor reported that the government, together with the Bank of England, were studying ways of improving the balance of

payments statistics, especially in the field of capital transactions. In the April 1957 *White Paper* an alternative presentation was introduced to distinguish between 'commercial' and 'monetary' transactions and the balancing item was distinguished in the main tables. In October 1957 the alternative presentation was incorporated in the main tables and the analysis was simplified.

Prices

In this field the *White Paper* made no recommendations, but much information was already available as both wholesale and retail monthly price indices were published before the war. Since the war these series have been reorganised. Based on 1948, new wholesale price indices were begun and publication of indices covering all stages of manufacture has now ceased. Instead, the Board of Trade publishes separately input, output, and consumption indices for various sectors of the economy, together with as many commodity series as are considered useful. Thus one overall indicator which Moore found satisfactory has been replaced by a number of series which are more economically meaningful. The intention to rebase the series on 1954 was announced in July 1956, and publication begins in 1958.

The retail price index has also been transformed since the war. The interim index, using largely pre-war expenditure information, replaced the old cost of living index in 1947. When the results of the 1953 expenditure enquiry became available a new index was started from January 1956. However, retail price movements lag behind those of wholesale and commodity prices and the latter are therefore more important in indicating future trends.

Consumption and Expenditure

The *White Paper* required monthly information on consumption but the only information available for a long time was the continuation of the pre-war voluntary series of retail sales. From 1946 these series were extended to include some independent retailers, and in 1952 the samples were revised with the aid of data from the census of distribution, but coverage was still uneven. In October 1955 comprehensive indices from January 1955 were introduced for independent retailers and in May 1956 for large-scale retailers. These now cover over one-half of total consumers' expenditure. Although the bias of voluntary returns remains, figures are weighted to give the right representation of different kinds of business and the sample was strengthened in October 1957. Retail sales was one of Moore's reliable indicators, but it lags rather than leads.

An important recent development in this field was the monthly enquiry into hire purchase sales and outstanding debts, started in October 1955. Before then little was known of this already committed element of expenditure, and thus of the impact of regulations upon it used since the war to control consumption. Information was still incomplete

however, as the enquiry did not cover credit sales agreements, which are an alternative to hire purchase. The statistics were extended to include such agreements from July 1956.

The Chancellor promised further information on personal consumption itself with an enquiry into expenditure and income of a sample of households. This had been recommended in 1951 by the Cost of Living Advisory Committee in order to keep up to date information available from the expenditure enquiry made in 1953, which formed the basis of the weights of the new retail price index. A continuous enquiry of 5 000 households was begun in January 1957. As yet, however, there is no information on consumers' intentions, a forward-looking series which has been available in the United States since 1946.

Incomes

Moore listed payrolls as a lagging series and corporate profits as a leading one, but data on both of these are scanty in this country.

Since 1941 the six-monthly earnings enquiries mentioned above have provided information on total earnings of a large section of operatives, but full results are subject to five months' delay (the key figures are published in four months). A monthly index is published of the average change in payment for a standard working week to all manual wage-earners, but since this is based partly on minimum rates it probably underestimates increases in wages when labour is in shorter supply. This index was revised in February 1957 to facilitate comparison with the new retail price index. Reliable information is available annually on total wages and salaries from the Inland Revenue, but the publication of these data takes nearly a year. Some information is also available from the census of production and other sources. The Chancellor proposed no improvement here, but the Inland Revenue has made investigations of the differences in timing between pay and tax receipts with a view to quarterly estimates of wages and salaries.

Improvements were promised, however, in the information on profits. This also comes from the Inland Revenue but is subject to considerably greater delay than that on wages and salaries because of the lengthy process of assessment of profits liable to tax. A voluntary sample return by a number of companies of quarterly estimates of profits has been started.[13] The first results of this enquiry related to 1954 but publication of results has to wait upon improvements in the representativeness of the sample.

Financial Statistics

The only figures listed in the 1944 *White Paper* which come under this heading are those of savings, on which 'regular' information was required. However, the savings item in the national income remains a residual and estimates can be made for only certain kinds of saving (e.g. building society shares and deposits). More information may be available in future from the expenditure and income enquiry.

However, this is useful in a national income analysis rather than as an indicator. Moore listed five financial indicators, three of which were leading series. Two of these were business failures and new incorporations, on which data are published by the Board of Trade but which are unlikely now to be such useful indicators of the future because they are affected by the complications of financial arrangements. The third leading indicator, stock exchange prices, is well provided for by a number of indices although the absence of information on the size of dealings means that weighting is arbitrary. The *Monthly Digest of Statistics* reprints the Actuaries Investment Index. Bank clearings and deposits —a coincident series—are also published but are not analysed by industry as are the figures for advances. The lagging series was of bank rates on business loans but this, even if available, would not be very useful as these rates have since the war been determined by government policy rather than by market forces. Yields on bonds and securities are more sensitive series, although these have been affected by dividend limitation policies. Two unofficial 'confidence' indices are also available, based on the difference in yield between ordinary shares and government bonds.

National Accounts

Combining information from many sources makes possible estimates of national income, production, and expenditure, and the gross national product was one of Moore's coincident series. Since 1941 annual estimates have been published of these aggregates. The amount of detail provided has steadily increased and since 1947 the emphasis has changed to a social accounting approach, analysing the relationships between different sectors of the economy. It is a measure of the recent improvements in official statistics that in September 1956 the London and Cambridge Economic Service published seasonally adjusted quarterly estimates of domestic production and expenditure at constant prices. In January 1957 the first official quarterly estimates were published of national expenditure. This publication was particularly welcomed by the London and Cambridge Economic Service since it provides to those outside government service new information on government revenue and expenditure. The official estimates are not seasonally adjusted and are in terms of current prices.

There remains for consideration the first proposal made by the Chancellor, which was to make all government statistics more up-to-date. Moore concluded that by the study of a number of leading indicators a turning point in economic activity may be detected more quickly, and the time lag will obviously be shorter the more quickly the figures are available. In this, one limiting factor is the speed with which completed forms are returned to government departments, and it seems[9, 10] that at present there is often considerable delay. Business men obviously stand to benefit from a government policy of maintaining a high level of

activity, and prompt statistical information is therefore in their own interests. Perhaps the most important next step is the improvement in public relations recommended by the Verdon Smith Committee.[9] Such an improvement would also help towards the improvement of the samples of firms making returns. Although few gaps remain in the list of overall indicators, most of the short-period enquiries are on a voluntary basis and the sample is not a fully representative one. If more firms are willing to co-operate a probability sample becomes practicable and the reliability of the results is improved.

Another difficulty which arises in the use of short-period indicators is the assessment of the influence of seasonal factors. Unless such an assessment can be made the significance of changes in the series cannot be appreciated. With the recently started series (e.g. of stocks and of capital expenditure) the seasonal variation cannot be calculated with any accuracy. But fortunately this problem will resolve itself as time goes on and the series should become increasingly useful as they are longer established.

My thanks are due to a number of civil servants who have been most helpful in answering my questions, and to Dr L. R. Klein for supplying references. The views expressed are, however, my own.

REFERENCES

[1] *Employment Policy* (1944). Cmd. 6527. HMSO, London.
[2] *Hansard*, 17th April 1956.
[3] *Hansard*, 1st August 1956.
[4] Moore, G. H. (1950). *Statistical Indicators of Cyclical Revivals and Recessions*, Occasional Paper No. 31, National Bureau of Economic Research.
[5] *Short-term Economic Forecasting* (1955). Studies in Income and Wealth, Volume 17, National Bureau of Economic Research.
[6] Economic Commission for Europe, Geneva (1955). 'Short-term business indicators in Western Europe', *Economic Bulletin for Europe*, **7**, No. 3, 34.
[7] Turner, H. A. (1955). 'Measuring unemployment', *J. R. Statist. Soc.*, A, **118**, 28.
[8] Bluestone, A. (1956). 'Overtime hours as an economic indicator', *Monthly Labour Review, Washington*, **79**, 1024.
[9] *Report of the Committee on the Censuses of Production and Distribution* (1954). Cmd. 9276. HMSO, London.
[10] Political and Economic Planning (1957). 'Statistics for government', *Planning*, **23**, No. 406.
[11] Dow, J. C. R. (1956). 'Surveys of investment intentions and other surveys of business intentions and expectations', Royal Statistical Society Study Section, Conference on Economic Statistics.
[12] *National Income Statistics. Sources and Methods* (1956). Studies in Official Statistics No. 3, HMSO, London.
[13] *Economic Trends* (1957). No. 43, HMSO, London.

PROBITS: AN APPLICATION TO OBSERVATIONAL DATA

JOHN K. FRIEND*
Operational Research Branch, BOAC

In this article Mr Friend first discusses the principle of probit analysis and the strong case that exists for its adoption in industry when a defective rate is being investigated for correlation or trend. He then describes how, in a certain investigation of this type in a steelworks, this intuitive case was borne out by a critical examination of the data and how certain obstacles, implicit in the application to observational data of a technique usually associated with experimental biology, were successfully overcome.

The Analysis of Quality

Where a statistical survey has to be based on observational data, simultaneous variations in many factors may make the analysis very much more complex than it would be if these factors could be made to vary one at a time under experimental conditions. A typical case is the problem faced by the industrial statistician when, in a manufacturing process, the quality of the article produced is affected by fluctuations in numerous contributory factors of production. In such circumstances the statistician is often called upon by management to provide answers to questions such as the following:

(1) Are there any significant time trends in quality?

(2) What measurable factors of production have a significant effect on quality?

(3) What change in quality could be expected from a given change in the level of any one of these factors?

If all the variables are normally distributed, the statistical analysis follows classical lines, based on the canons of time series analysis, analysis of variance, and regression theory (multiple regression if the contributory factors of production are inter-related). If some of the variables are non-normal, a mathematical transformation may suffice to make the theory applicable.

Defective Rates as Measures of Quality

'Quality' is a term which needs definition with respect to the investigation concerned. It will normally be judged by the measurement of some specific characteristic of the article produced, such as its weight, its life under stated conditions, or the percentage content of some constituent. However, in some industrial processes control of quality takes

* Formerly with United Steel Companies Ltd.

the form not of measuring the quality of each article, or even of samples from the production line, but of testing all the articles, or samples of them, to ensure that they conform to certain minimum standards; this might involve the passing of the articles through a gauge, or a visual inspection for defects, or a mechanical test of strength. In this case the only direct measure of fluctuations in quality would be the failure rate of the articles submitted to the test. Following normal industrial terminology, this rate will be known as the defective rate and those articles which fail will be known as defectives, although the work that follows may well be applicable to other forms of response rates in other than industrial fields.

For single articles, the defective rate is limited to the two values of zero and 100%; for small batches of articles its value as an index of quality is limited by the fact that defectives occur only in integral numbers and, assuming defectives are the exception rather than the rule, the frequency of a zero defective rate will be high. It is only between large batches, then, that a defective rate becomes a sensitive indicator of changes in quality, and it will often be observed that even between large batches its distribution is markedly skew and that it cannot be treated as a normal variable. The following sections show how a transformation can be applied which may be expected to normalise the distribution.

Defining a Quality Index

Failure or success in the passing of a test often depends ultimately on one measurable factor; whether or not a steel rod passes through a gauge depends directly on its diameter, and whether or not a man at A catches a train at B depends on the speed at which he can run. In other tests, no obvious measurable factor may govern success, but the existence of such a factor can be readily visualised.

The statistical theory of biological assay is a case in point. A certain concentration of insecticide is applied to a group of insects; the only visible effect is that some of the insects die whereas others survive. It may, however, be assumed that each insect has a certain 'tolerance index' to the poison, based on many internal and external factors, and that an insect will die only if its 'tolerance index' lies below a certain limiting value. The choice of a scale for this index is arbitrary, but it is found that the assumption that such an index exists, and that it is normally distributed between insects, provides a mathematical model which fits the observed facts closely and produces valuable practical results. This is the basic idea of probit analysis, which has proved invaluable in many branches of experimental science.

In the same way as an index of 'tolerance' is formulated in biological assay, an index of quality can often be formulated for manufactured articles submitted to a test. 'Quality' may be imagined as a continuous variable, varying normally between articles, such that any articles whose quality is less than a given value will fail in the test. 'Quality'

will not usually be hard to visualise in this way; however, before such an index can be handled numerically, a scale must be chosen.

Choice of Scale

The most important point on the scale of the quality index will be the critical point between success and failure in the test. This is chosen as the zero of the scale, so that an article with positive index will pass, and one with negative index will fail. (In standard probit analysis, the critical point is taken as 5, to avoid the occurrence of negative values; in this work, however, 0 is preferable, as it avoids certain arithmetical complications.)

The remainder of the scale is fixed by defining a 'standard batch' of articles, and taking the standard deviation of the quality index within

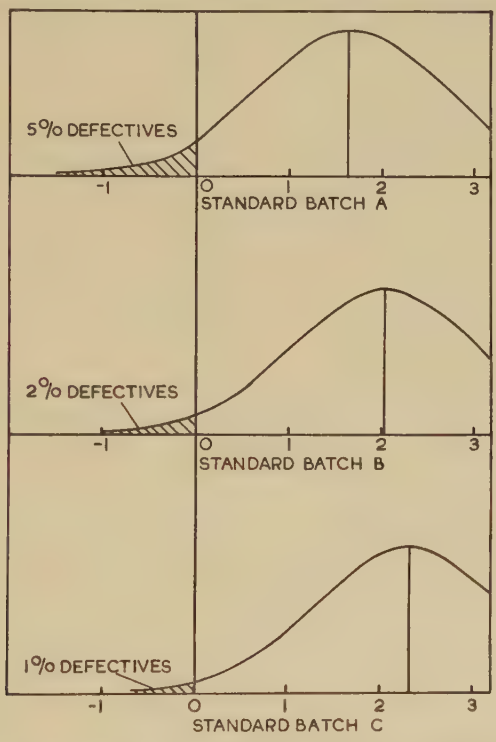


FIG. 1. Percentage of defectives per standard batch for three batches of different mean quality indices, shown on the horizontal scales.

this batch as the unit of the scale. The 'standard batch' may be any natural unit of production; if the raw materials for a number of articles are received or processed at one time, or if the articles to be tested are sampled in batches drawn from a larger population, choice of standard batch is clear. Otherwise daily, weekly, or monthly production totals may be chosen. The use of a standard batch assumes that the variation in quality within each standard batch is the same (although the average quality may vary from batch to batch), and it is important that the

data should be examined critically to test this assumption. It will usually be reasonable provided that none of the factors of production are considerably more erratic at some times than at others.

Fig. 1 shows how, when the proportion of defectives in any one standard batch is known, its mean quality index can be estimated. The distribution of quality inside a standard batch is represented by a normal frequency curve with unit variance; if this curve is gradually shifted along the scale so that its mean varies from $-\infty$ to $+\infty$, the percentage of defectives decreases from 100% to zero. Corresponding to any given percentage of defectives, there is therefore one position of the mean. Fig. 1 shows the positions corresponding to 5%, 2%, and 1%; the means of these curves are 1·64, 2·05, and 2·33 respectively. This transformation of the percentages is equivalent to the probit transformation, and the value thus found will be known as the 'zero-based probit' (to distinguish it from the usual probit, which is based on an origin of 5 instead of 0). Hence for any standard batch, an index of mean quality can be estimated by consulting tables of the normal integral or probit tables (tables of the normal integral must be used inversely, whereas probit tables are read directly but 5 must be subtracted from the tabulated figure; the percentage of non-defectives should be used in place of the percentage of defectives to give the correct sign to the probit). For a non-standard batch, provided that an estimate of its internal variance in quality, σ^2, is known, the mean batch quality can be estimated by multiplying the zero-based probit of the overall defective rate by σ.

In practice, as a standard batch contains only a finite number of articles, the distribution of quality in the batch will only approximate to the normal form, and the calculated batch quality will only be an estimate of the actual batch quality. The larger the standard batch, the more accurate the estimate. It can be particularly awkward if a batch contains no defectives (or, much less likely, 100% defectives) in which case the probit transformation will give no finite value for the average quality index. This obviously need not mean infinitely good (or infinitely bad) quality. It merely means that the standard batch is not large enough to show the one inevitable defective (or, in the 100% case, non-defective) which, on the assumption of normality of the quality distribution, can be expected to occur sooner or later whatever the batch quality. It is advisable, therefore, to choose as standard batch a unit large enough to make this occurrence rare.

Advantages of the Probit Transformation

The transformation of the defective rate to a probit quality index, based on a suitably chosen standard batch, may thus be expected to yield a more basic variable, whose variation between standard batches may be expected to be normal in form. It is possible that other transformations are just as effective as (perhaps in some cases more effective than) the probit transformation in achieving normality, but the

concept of some hidden basic characteristic of the article produced being responsible for success or failure in the test gives a strong intuitive case for choice of the probit transformation. A second important point in favour of this transformation is that not only does it allow analyses to be performed between standard batches and other large batches where the transformation can be accurately applied, but, as will be shown later, the technique of fitting a probit regression line allows the analysis to be extended to variations between smaller batches or between individuals where, as previously described, the defective rate is a crude indicator of variations in quality, and where the frequent occurrence of zero defective rates makes it impossible to carry out any normalising transformation directly.

In this consideration of the advantages of the probit transformation, it should at the same time be stressed that, in investigations which are apparently amenable to its use, a critical examination of the data should first be made to ensure that the basic assumptions are satisfied —in practice it may sometimes be found that the defectives are not of a type stemming from fluctuations in one basic property, or that the variance within batches is not constant. However, where the probit method is applicable, its use can often lead to results of vital importance; on the following pages I shall attempt to demonstrate this by describing an application in the steelmaking field.

A Steelmaking Application: Outline of Process

The idea of a quality index based on the probit transformation first suggested itself during an investigation into the occurrence of defectives in the manufacture of a certain rolled steel product at a branch of the United Steel Companies Ltd. The process of manufacture is briefly as follows.

First, a 'cast' of steel is made in the melting shop, and samples from the cast are analysed to confirm that the steel falls within required limits as regards silicon, carbon, manganese, sulphur, and phosphorus composition. The molten cast is then poured into ingots which, after solidifying and being consigned to a soaking pit to distribute the heat evenly, are rolled into their final form, each cast being sufficient for approximately twenty lengths of the finished product. Each length is then put through a certain test, those which fail being rejected as defectives.

Every few weeks a 'rolling' of this product is made, lasting three or four days and consisting of 50 to 100 casts, after which production is switched to other types of steel. Thus the natural units into which production falls are primarily the cast and, on a larger scale, the rolling. For these units the percentage of defective lengths, termed the defective rate, serves as a basis for the assessment of quality.

Variation between Casts

The very nature of the process, with mature judgement in the melting

of the cast largely dictating the quality of the steel, made it obvious that the quality of an individual length would be largely dependent on the quality of the cast.

The importance of the cast is illustrated in Table I, which shows the distribution of casts according to the number of defectives for four typical rollings. For all rollings where there were sufficient defectives, such as A, B, and C, a Poisson distribution gave a poor fit, whereas a negative binomial distribution (i.e. Poisson with a variable mean) fitted closely in every case, giving evidence of considerable variability between casts in their proneness to the defect.

TABLE I

Distributions of number of defectives per cast

Rolling	Frequency of Casts having Number of Defectives Equal to:										Total No. of Casts
	0	1	2	3	4	5	6	7	8	9	
A	48	20	6	7	3	1	1	1	1	1	89
B	74	24	10	2	1	1	—	—	—	—	112
C	96	15	2	1	1	—	—	—	—	—	115
D	48	2	1	—	—	—	—	—	—	—	51

This suggested that a fruitful field for analysis would be the seven factors of production which were measured for each cast. Of these, three were found to exert a significant influence on the failure rate: the manganese content, the silicon content, and the track time (the time elapsed from pouring the molten cast into ingots to the insertion in the soaking pits). The significance test used in each case was a simple t-test between defective-free and other casts.

The way in which the defective rate was found to depend on one of these variables, silicon, is shown in Fig. 2(a). A number of rollings, chosen to include both high and low defective rates and high and low silicon levels, were each split into four approximately equal groups of casts according to silicon content, and in each group the proportion of defectives was plotted against the mean silicon. Thus, for each rolling, a rough indication of the effect of silicon on the defective rate was obtained. It can be seen from Fig. 2(a) that the higher the average level of the defective rate, the more pronounced becomes the effect of silicon.

Fig. 2(b) shows the result of applying a simple probit transformation to the defective rate. All the lines are now parallel, giving strong evidence that a simple linear relationship between silicon and a more basic variable 'quality' lies behind the more complex relationship between silicon and the defective rate, and thus providing observational as well as intuitive justification for basing an analysis of quality on the probit transformation.

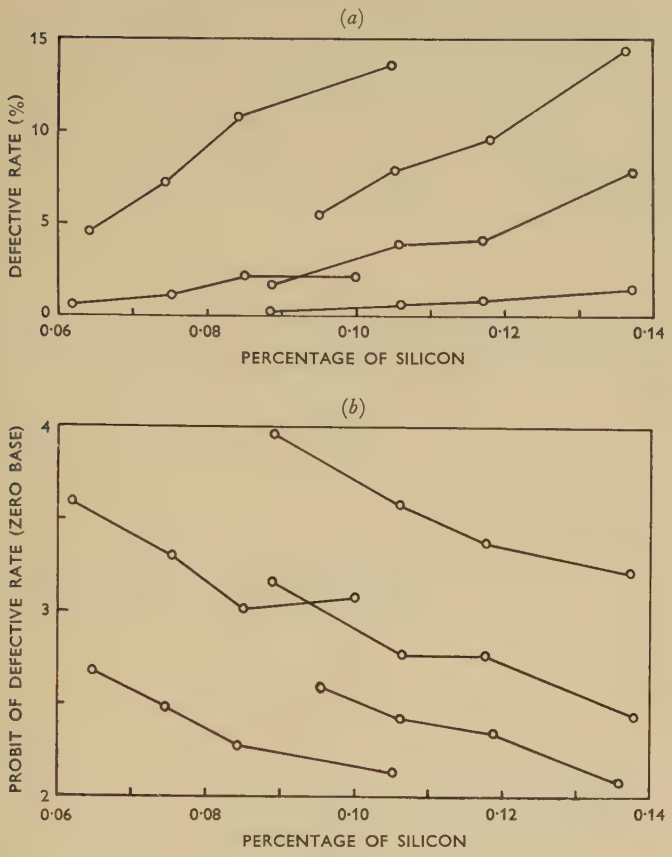


FIG. 2. Effect of silicon on defective rate for five typical rollings.

The Rolling as Standard Batch

To define a quality index, all that is required is the choice of a unit as standard batch in order to fix the scale. In this case the cast proved unsuitable, despite its fundamental importance; Table I shows that a majority of casts contain no defectives, and therefore do not permit the estimation of finite quality indices. This suggested the rolling as standard batch, and this proved a suitable choice as, with 1000–2000 lengths per rolling, a completely defective-free rolling was rare; also the rolling appeared to satisfy the condition that, in all relevant factors, the degree of variation did not change from one rolling to the next. Choice of rolling as standard batch allowed a mean quality index to be estimated for each rolling, as in Fig. 1, and a complete analysis to be carried out of variations in quality from rolling to rolling.

In Fig. 3(*a*) the percentage of defectives is plotted for 72 consecutive rollings. The variation from rolling to rolling was much greater than chance alone could account for; the general trend appeared to be one of improvement, but occasionally disturbing peaks occurred when the rate was abnormally high.

The mean quality index estimated from each defective rate is shown in Fig. 3(*b*). (For the five rollings with no defectives a nominal index corresponding to half a defective per rolling was taken.) It will be noticed that the scatter is much more even than in Fig. 3(*a*), and it

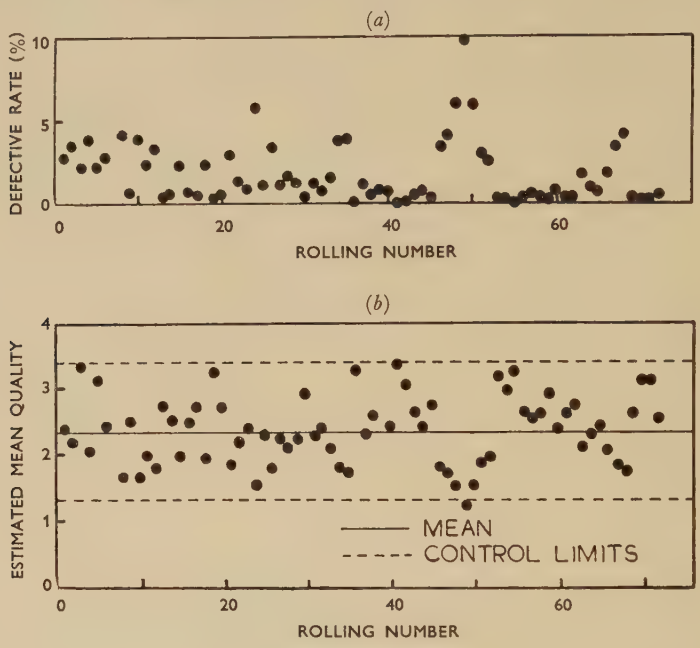


Fig. 3. Defective rate and estimated mean quality for seventy-two rollings.

appears that the peaks in the defective rate are due not to any freak conditions but to perfectly normal minima in the underlying fluctuations in quality. A regression analysis showed that no significant linear or quadratic trends were present, but an autocorrelation analysis verified that variation was not entirely random. One in forty (2σ) control limits were calculated, and these have been superimposed on Fig. 3(*b*). This allowed a check to be maintained on subsequent fluctuations in the level of quality.

Assessing the Influence of One Factor

Potentially, one of the most promising ways of reducing the incidence of defectives lay in changing the average level of one or more of the three cast factors known to affect cast quality; silicon, manganese, and track time. Before some economic assessment of the consequences of such a step could be made, however, a reliable estimate of the effect on the defective rate of any specified reductions in the three factors was required. This could not be obtained from a correlation of rolling averages, as in normal circumstances the levels of the factors varied little from rolling to rolling and were evidently not associated with the long-term fluctuations in quality. This left only the possibility of an

analysis of variations from cast to cast within a rolling or group of rollings, and Fig. 2(*b*) gave an indication of how this might be carried out. If, as in Fig. 2(*b*), the casts in a rolling or group of rollings are split into subgroups according to the value of the factor X whose influence is being investigated and a plot is made of the probit of the defective rate against X, then if a straight line is fitted with slope B, the regression coefficient b of quality on X will be related to B by the equation

$$b = B\sigma$$

where σ is the standard deviation of quality within subgroups. This relationship arises from the multiplication of the probit by σ to obtain the estimated mean quality of a subgroup. However, σ^2 is equal to the overall between-and-within cast variance σ_T^2 of the rolling or group of rollings, less the proportion due to between-cast variation in X, that is:

$$\sigma^2 = \sigma_T^2 - b^2\sigma_X^2$$

where σ_X^2 is the total between-cast variance in X.

These two equations combine to give

$$b = B\sqrt{\left(\frac{\sigma_T^2}{1 + B^2\sigma_X^2}\right)}$$

For a single rolling $\sigma_T^2 = 1$, and for a group of rollings $\sigma_T^2 = 1 + \sigma_R^2$ where σ_R^2, the between-rolling variance in quality, can be estimated by the mean square deviation of the estimated quality indices of the rollings concerned, weighted according to the number of casts per rolling. Also, σ_X^2 can be estimated by ordinary mean-square methods, so all that remains to evaluate b is to find a satisfactory estimate of B. This can be achieved by fitting a regression line by eye and then using successive approximations to reach the maximum-likelihood estimate of the line. This is a standard technique of probit analysis and is described by Finney[1] and in the prefaces to the Tables of Fisher and Yates[2] and Pearson[3]. How the method can be used in practice, with no Tables beyond those of the normal distribution, is explained by an example below.

Fitting a Probit Regression Line

In Table II, 240 casts, representing the output of three consecutive rollings, are grouped according to percentage silicon content, the overall defective rate within each group and the corresponding zero-based probit being shown.

The probit is plotted against silicon in Fig. 4 with the number of casts contributing to each point indicated in brackets. The tendency for quality to decrease with increasing silicon is clear; as a first estimate of the relationship, a straight line was fitted by eye, giving a regression coefficient of $B = -13{\cdot}9$.

TABLE II

Effect of percentage of silicon on quality

Silicon %	No. of Casts	Defectives %	Probit (Zero-base)
0·070–0·079	6	2·52	1·96
0·080–0·089	8	3·76	1·78
0·090–0·099	20	1·68	2·12
0·100–0·109	79	4·24	1·72
0·110–0·119	65	5·01	1·64
0·120–0·129	34	8·67	1·36
0·130–0·139	19	6·28	1·53
0·140–0·149	6	6·14	1·54
0·150–0·159	3	21·96	0·77

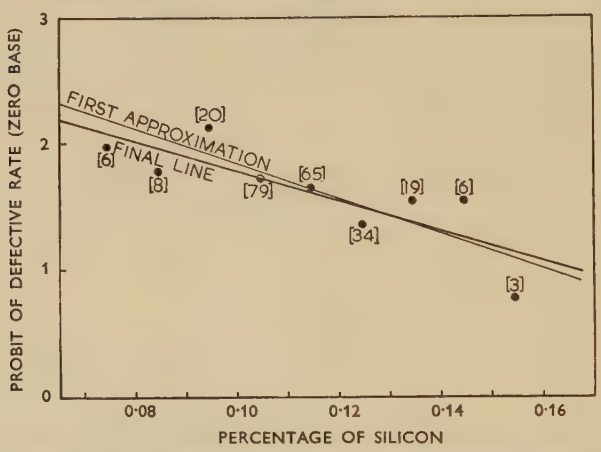


FIG. 4. Zero-based probit of defective rate plotted against silicon content; the number of casts represented by each point is shown in brackets.

Table III shows the process by which this first line was used to give a second approximation to the maximum-likelihood estimate.

The first four columns repeat the initial data of Table II; x is the mean silicon content of each group in simplified working units, n the number of casts, p the proportion of non-defective lengths, and y the corresponding probit. The remainder of the table is based on the line fitted by eye. Y is the probit read off the line for each x, P is the corresponding proportion of non-defectives, i.e. the normal integral corresponding to Y, and Z is the corresponding normal ordinate. Both P and Z are obtainable direct from tables. Then a weighting factor $w = Z^2/P(1-P)$ is calculated, which converts each frequency to a weighted frequency nw. The correction $(p-P)/Z$ is then added to Y to give y', the 'working probit.' The second approximation to the regression line is then calculated by a simple regression of y' on x, with group frequencies nw, that is

$$Y = A + Bx$$

where $$B = \frac{\Sigma nwxy' - \Sigma nwy'\Sigma nwx/\Sigma nw}{\Sigma nwx^2 - (\Sigma nwx)^2/\Sigma nw} \quad \text{and} \quad A = \frac{\Sigma nwy'}{\Sigma nw} - B\frac{\Sigma nwx}{\Sigma nw}$$

TABLE III

Fitting a probit regression line

x	n	p	y	Y	P	Z	$w = \frac{Z^2}{P(1-P)}$	nw	$\frac{p-P}{Z}$	y'
−0·04	6	0·974 8	1·96	2·19	0·987 5	0·036 3	0·093 5	0·56	−0·30	1·89
−0·03	8	0·962 4	1·78	2·05	0·979 8	0·048 8	0·120 4	0·96	−0·36	1·69
−0·02	20	0·983 2	2·12	1·91	0·971 9	0·064 4	0·151 9	3·04	+0·18	2·09
−0·01	79	0·957 6	1·72	1·78	0·962 5	0·081 8	0·185 3	14·64	−0·06	1·72
0	65	0·949 9	1·64	1·64	0·949 5	0·104 0	0·225 4	14·65	0·00	1·64
+0·01	34	0·913 3	1·36	1·50	0·933 2	0·129 5	0·269 1	9·15	−0·15	1·35
+0·02	19	0·937 2	1·53	1·36	0·913 1	0·158 1	0·315 5	5·99	+0·10	1·46
+0·03	6	0·938 6	1·54	1·22	0·888 0	0·189 5	0·363 4	2·18	+0·26	1·48
+0·04	3	0·780 4	0·77	1·08	0·859 9	0·222 7	0·411 6	1·23	−0·36	0·72

If this line differs substantially from the first approximation the process can be repeated; several iterations may sometimes be necessary if a high degree of accuracy is required. In this case the initial visual estimate of B was $-13{\cdot}9$ and successive approximations gave $-12{\cdot}7$, $-11{\cdot}7$, $-12{\cdot}1$, and $-12{\cdot}1$, at which stage $-12{\cdot}1$ was accepted as the best estimate.

Estimation of the Components of Variance

The values of B for silicon, manganese, and track time were estimated in this way from a large group of 52 rollings containing 3 832 casts and, the values of σ_T^2 and of each σ_X^2 having been estimated by normal mean-square methods, the regression coefficients b of quality on each variable were obtained. However, two of these variables, silicon and manganese, proved to be significantly intercorrelated, and so it was decided that, to estimate the independent effect of each variable on quality, this intercorrelation would have to be taken into account and partial regression coefficients obtained. This involved estimation not only of the intercorrelation coefficients between silicon, manganese, and track time, and the between-cast variance of each of these variables, but also of the between-cast variance in quality. Estimation of this latter quantity by mean-square methods was not possible, and so a graphical approach was adopted, as described below.

For any group of rollings, the overall variance in quality can be visualised in three components:

σ_R^2 = variance between rollings

σ_C^2 = variance between casts within rollings

σ_L^2 = variance between lengths within casts

Of these, σ_R^2 can be readily estimated as discussed before, by calculating the mean square deviation of the estimated mean quality indices of the individual rollings, each being weighted according to the number of lengths in the rolling. Owing to the non-randomness of mean rolling quality, apparent in Fig. 3(b), σ_R^2 tends to be less for a smaller group of rollings than for a larger one. For the 52 rollings concerned, its value was found to be 0·40.

It was deemed a justifiable assumption that the other two components, σ_C^2 and σ_L^2, were constant for all rollings; their sum had already been assumed constant by assuming that each rolling is a 'standard batch' with the same unit variance. It was found that if the distribution of k, the probit of the cast defective rate, was known over a group of rollings, σ_C^2 and σ_L^2 could be estimated by the formulae:

$$\sigma_C^2 = \frac{\sigma_k^2 - \sigma_R^2 - (2 + \bar{k}^2)/2n}{1 + \sigma_k^2}$$

and

$$\sigma_L^2 = 1 - \sigma_C^2 = \frac{1 + \sigma_R^2 + (2 + \bar{k}^2)/2n}{1 + \sigma_k^2}$$

where n ($= 20$) is the number of lengths per cast and $\bar{k}$ and σ_k^2 are the mean and variance of k. The derivation of these formulae is explained in the Appendix.

Table IV shows the distribution of defectives per cast for the group of 52 rollings.

TABLE IV

Distribution of defectives and of 'k'; 52 rollings

d	No. of Casts with *d* Defectives	Percentage of Casts with over *d* Defectives	Defective Rate %	*k* = Probit of Defective Rate
0	2 692	29·75	0	+ ∞
1	537	15·74	5	+1·645
2	236	9·58	10	+1·282
3	131	6·16	15	+1·036
4	81	4·04	20	+0·842
5	55	2·63	25	+0·674
6	27	1·91	30	+0·524
7	32	1·07	35	+0·385
8	15	0·68	40	+0·253
9	10	0·42	45	+0·126
10	9	0·183	50	0·000
11	2	0·130 5	55	−0·126
12	3	0·052 5	60	−0·253
13	1	0·026 3	65	−0·385
14	1	0·000 0	70	−0·524
Total	3 832			

With 20 lengths per cast, incidence rates of 0, 1, 2, 3, etc., defectives per cast correspond to defective rates of 0%, 5%, 10%, etc. These percentages were transformed to probits, k, and Fig. 5 shows on a probability grid the cumulative frequencies of the different values of k. The distribution is 'stepped', as only an integral number of lengths can

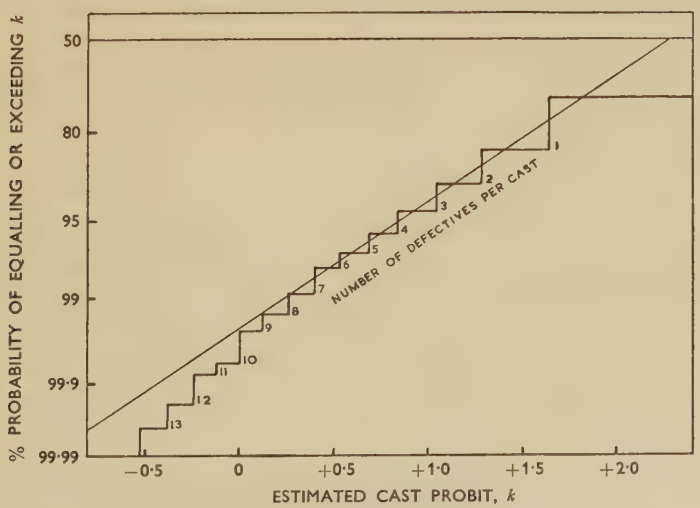


Fig. 5. Cumulative distribution of estimated cast probit, k, for fifty-two rollings.

fail in each cast of twenty; however, had the cast size been variable the distribution would have been more continuous.

On Fig. 5, a line of best fit has been drawn visually to estimate the distribution of k. Owing to the width of the 'steps' for large k, and the small sample size for small k, the most accurate part of the distribution for fitting was in the region $0{\cdot}5 < k < 1{\cdot}5$. In this region it can be seen that a straight line gives an extremely good fit; and it should be noted that even for $k < 0{\cdot}5$, where the actual distribution appears to diverge, small sample sizes make the divergence not significant. The intercept of the line on the 50% level gives the mean, $\bar{k}$, and the difference between this and the intercept on the 84·1% level gives σ_k. Their values were estimated at 2·28 and 0·90 respectively and, substituting in the formulae, σ_C^2 and σ_L^2 were estimated at 0·142 and 0·858 respectively, i.e. $\sigma_C = 0{\cdot}377$ and $\sigma_L = 0{\cdot}926$.

The accuracy of this graphical method of estimation was borne out by the almost identical values obtained when the 52 rollings were split into four separate groups of 13, and the estimation carried out on each group separately.

Conclusion of the Analysis

The estimation of σ_C^2 removed the last remaining obstacle to the evaluation of partial regression coefficients. Figures of $-9{\cdot}5$, $-1{\cdot}74$ and $-0{\cdot}0201$ were obtained for silicon, manganese, and track time respectively.

The three partial regression coefficients were presented to management in the form of a number of parallel scales, as shown in Fig. 6. This is not a nomogram, but something simpler; the scales cover the full range of variation in each factor, and the silicon, manganese, and track time scales are so related to the quality index scale that equal distances represent equivalent changes. Shown flanking the quality scale are two scales for the defective rate designed to answer the questions:

1. What effect would a given change in the silicon, manganese, or track time level be expected to have in reducing the defective rate in a given rolling?
2. What effect would such a change be expected to have in reducing the overall rate for a group of rollings with $\sigma_R^2 = 0{\cdot}40$?

These two questions are designed to test the efficacy of a change in the silicon, manganese, or track time level (1) as an immediate remedy for a sudden increase in the defective rate, and (2) as a long-term means of reducing the incidence of defectives. As an example, suppose it is technically possible to reduce the average level of silicon percentage from 0·100 to 0·080. The reduction of 0·020 can be transferred from the silicon scale to any point on the quality scale, giving an increase of 0·19 ($= -9{\cdot}5 \times -0{\cdot}02$, applying the partial regression coefficient) in mean quality, and a decrease in the defective rate, the amount of

which depends on its initial level. For instance, using the left-hand scale, in a rolling with 5% defectives the rate would be reduced to approximately 3·3% while a rate of 0·5% would be reduced to approximately 0·29%. An example of the use of the right-hand scale to predict long term changes is given by considering the 52 rollings of Table IV, whose overall defective rate was 3·7%. If it was decided, after these rollings, to effect a permanent reduction of 0·020 in the silicon level, the overall level of the defective rate could be expected to drop to 2·2% in subsequent rollings.

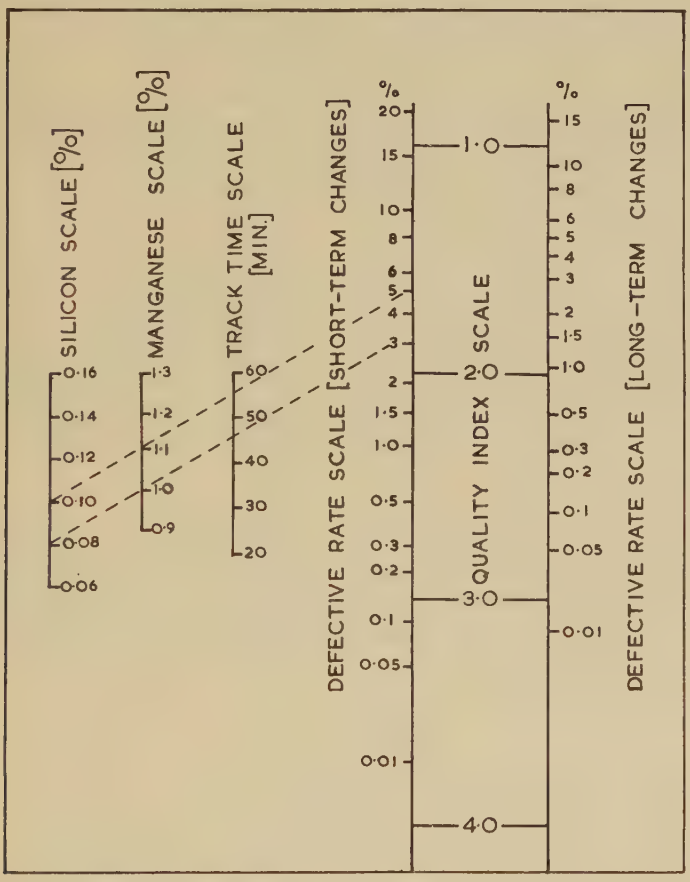


FIG. 6. Chart showing equivalent changes in quality index and defective rate to be expected from changes in manufacturing conditions. *Example:* for a rolling with 5% defectives and a silicon content of 0·1%, a reduction of silicon to 0·08% would be expected to reduce the defective rate to 3·3%.

So, by referring from the silicon, manganese, or track time scale direct to the appropriate defective rate scale, a quick decision can be taken by management on whether a projected decrease in one of the factors is worth while as a means of reducing the defective rate, or, conversely, whether an increase in one of the factors, contemplated for other reasons, would be unduly detrimental to the defective rate.

General Conclusions

The above analysis gives an example of the way in which the probit technique may be brought into play in fields where data cannot be obtained by experimentation with controlled samples, but must be based on the observation of a system which may consist of a complex structure of groupings of different degrees of homogeneity, influenced by simultaneous variations in a multiplicity of factors. In such a system, it has been shown that the probit technique can be effective in the analysis of a response rate or defective rate, provided that the structure of the system is examined first to ensure that certain conditions are satisfied; in particular, the existence of some large units suitable for use as 'standard batch' must be established.

The use of the probit transformation is, of course, a weapon much more adaptable to estimation problems than to significance testing and the setting of confidence limits; it will be noticed, in the above analysis, that such significance tests as were carried out involved the defective rate rather than the quality index.

Although in this and other ways the theory is in no way complete or rigorous, it not only produced valuable results in the situation described, but offered a fresh standpoint from which to view the defective problem. It is hoped that this account of the analysis will give some impetus to the wider application of a technique, long standard in biological statistics, not only to industrial problems but also to other fields where reliance must be placed on observation rather than experimentation.

Acknowledgements

I wish to thank Mr F. H. Saniter, Director of the Research and Development Department, the United Steel Companies Limited, for permission to publish this article, and the management of the branch concerned for permission to quote the steelmaking application.

APPENDIX

Derivation of Formulae for σ_C^2 and σ_L^2

In a group of rollings with mean quality M and between-rolling variance σ_R^2, let the quality of the jth length of the ith cast be

$$q_{ij} = M + C_i + l_{ij}$$

where C_i (mean zero, variance $\sigma_C^2 + \sigma_R^2$) depends on cast properties and l_{ij} (mean zero, variance σ_L^2) on the properties of the individual length.

If there are n lengths in the ith cast, they may be considered as a sample of n from the population of mean $q_i = M + C_i$ and variance σ_L^2. The mean index of the sample will be:

$$\bar{q}_i = M + C_i + \frac{1}{n}\Sigma_j l_{ij}$$

and the within-sample variance will be:

$$S_i^2 = \frac{1}{n-1}\left[\Sigma l_{ij}^2 - \frac{1}{n}(\Sigma l_{ij})^2\right]$$

Let the expected proportion of defectives in this sample, given $\bar{q}_i$ and S_i^2, be p_i. Then, if k_i is the probit of p_i, $k_i = \bar{q}_i/S_i$.

From approximate formulae for the mean and variance of a quotient of two independent variates, the mean and variance of k_i over all casts are given by

$$\bar{k} = \frac{M}{\sigma_L}$$

and

$$\sigma_k^2 = \frac{1}{\sigma_L^2}\operatorname{var}(\bar{q}_i) + \frac{M^2}{\sigma_L^4}\operatorname{var}(S_i)$$

$$= \frac{1}{\sigma_L^2}\left(\sigma_C^2 + \sigma_R^2 + \frac{\sigma_L^2}{n}\right) + \frac{M^2}{\sigma_L^4}\cdot\frac{\sigma_L^2}{2n}$$

$$= \frac{\sigma_C^2 + \sigma_R^2}{\sigma_L^2} + \frac{1}{n} + \frac{\bar{k}^2}{2n}$$

Also, by definition of the standard batch, $\sigma_C^2 + \sigma_L^2 = 1$, and substitution of this relationship quickly gives the required formulae for σ_C^2 and σ_L^2.

REFERENCES

[1] Finney, D. J. (1952). *Probit Analysis.* Cambridge University Press.

[2] Fisher, R. A. and Yates, F. (1953). *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, Edinburgh and London.

[3] Pearson, E. S. and Hartley, H. O. (Editors) (1954). *Biometrika Tables for Statisticians, Vol. I.* Cambridge University Press.

STATISTICS OF AN AUCTION SALE

DAVID C. DUNCAN

In this article Mr Duncan presents facts and figures relating to a sheep auction, which should be of interest to economists studying market situations and may serve as an incentive to further investigations along similar lines.

Economics is remarkable in that whereas it sets out to study the hard facts of man's necessity to work and make a living with a limited amount of time and energy at his disposal, it begins by studying market situations which are seldom, if ever, found in everyday life. Every student of economics has heard of the perfect market, but what economist has ever studied one in operation? Do any exist? The expositors of economics if approached in this way would have their justifications ready, since this is an old skeleton whose bones are well picked. They would readily admit that these are unreal situations, hedged around with so many qualifications that in real life they do not occur. But this, they would argue, does not deprive them of their value. To learn to understand the real market situation, which is very complex, it is necessary to start by studying simple cases, 'ideal types', and the simplest of these is the perfect market. In fairness to the economists it should be mentioned that this happens in psychology too. In order to understand the more easily the complexities of human behaviour, students of psychology are bidden by some psychologists to study our simpler cousins, the rats, dogs, sheep, and chimpanzees, as 'pedagogical paradigms', and naturally what by one discipline is considered sound common sense sounds like arrant nonsense to the other.

Industrial psychologists on the whole prefer to approach their subject in a more inductive manner, starting by studying the simplest situations they can find in real life, and in time attempting to build up theories on their observations of these situations. Having at heart the interests of economists as well as psychologists, I decided to try to have the best of both worlds, and help both economists and psychologists, by studying a real life situation approximating to a perfect market and at the same time having some interest for the psychologist studying primitive situations. This involved a search for a real life market situation which fulfilled as far as possible the following conditions:

(*a*) A homogeneous commodity;

(*b*) Large numbers of buyers and sellers;

(*c*) Perfect information among buyers and sellers on price changes;

(*d*) Pairing of buyer and seller by chance and not by preference.

The quest for a perfect market was, of course, vain, and one might also consider that it would not have been a very interesting subject of study even if we had found it. We did, however, find an interesting approximation which is the true subject of this paper—a one-day auction sale which took place in the Scottish Borders in August 1955. It involved 14 319 half-bred ewe lambs which were sold in 215 lots of approximately 50 and 100.

Taking first the economist's interests, one could argue that the sheep were a homogeneous commodity, since all were lambs of the same breed, sex, and age. The situation is, however, complicated by the effect on the size and quality of the sheep of the parent stock, the altitude and fertility of the farm of origin, and the methods of husbandry practised by the farmer. The most that can be said to offset this complication is that the range of quality is open to inspection, and that since this is a matter of vital statistics, the commodity could be said to vary in a known fashion. The statistician's prediction of the result would be a normal distribution of the number of sheep and of lots over a range of prices. The second condition of the perfect market was easily fulfilled in that there were large numbers of buyers and sellers. It could also be argued that perfect market conditions were also satisfied in that, except for perhaps the first twenty lots, all buyers could see what prices were being paid for what quality, and in what way they varied. The order of selling of lots was decided by ballot, so that the distribution of large and small lots, and of good-quality and poor-quality lots, was random over the period of the sale.

The pairing of buyer and seller was not investigated at the time of the sale. It is likely, however, that a fair amount of pairing by preference took place, but associated with the range of quality offered. On the one hand buyers would have the opportunity of inspecting the lots before they arrived in the sale ring, and could at the auction buy on the evidence of their eyes and hands; but on the other hand many buyers would know before the lots entered the ring, and without having seen the sheep, what sort of sheep to expect from a given farm. In addition, since buyers and sellers would generally be members of the same community in the business of agriculture, many of the buyer-seller pairs would be personally acquainted. This implies, however, that buyers from a well-known farmer might see that he obtained a 'just price' for his sheep rather than that they would pay for his sheep more than they were worth. That is, the pairing of buyers and sellers is made likely only because the buyers are in possession of such complete information about the nature of the commodity. Were all the buyers strangers to the district, pairing might be by chance, but the information possessed would be incomplete. From the specification of this approximation to the perfect market we turn to the more statistical aspects of the study.

In the first place, the statistician's prediction would be fulfilled in that the data provided an approximately normal distribution of lots

and sheep over a range of prices. These distributions are given in Table I. The average price of a sheep for both distributions was 197 shillings, the standard deviation was 23 shillings, and the mode was 200 shillings;

TABLE I

Distribution of sales over price range

Price (shillings per head)	112 to 130	132 to 150	152 to 160	162 to 170	172 to 180	182 to 190
No. of sheep	80	250	685	945	1 540	1 980
No. of lots	2	5	10	15	27	30

Price (shillings per head)	192 to 200	202 to 210	212 to 220	222 to 230	232 to 240	242 to 260
No. of sheep	2 767	2 410	1 490	1 010	842	320
No. of lots	36	37	21	16	11	5

the median of the distribution by sheep was 200 shillings. The distributions therefore had a slight negative skewness, the 'tail' being longer at the lower end of the price range. There was a tendency for bigger lots of sheep to gain a slightly higher price, which is illustrated in Table II by the relative positions of the quartiles for lots and sheep.

TABLE II

Quartile prices (shillings per head)

	Lower Quartile	Median	Upper Quartile
Lots	180	198	210
Sheep	182	200	212

This tendency, however, was not very marked, as both large and small lots were found throughout the price range. In particular, the lots gaining the top prices were below average in size. The distribution of prices was also affected by other buyer preferences. All bids were made in units of two shillings, and buyers showed a significant preference for stopping bidding at the 'round pound'. As a result, sales at the 'round pound' accounted for 20·5 per cent of the total sales, as against an expected value of about 10 per cent. This preference is significant even when the modal value of 200 shillings, at which price 7·7 per cent of the sheep were sold, is excluded from consideration. However, in the later parts of the sale, when bidding was weaker, a bold bid two shillings past the pound sometimes carried the sale, and as a result there was a modal frequency at 182 shillings at which 5·6

per cent of the sheep were sold, the expected frequency being about 1·3 per cent if the whole range is equally subdivided, and about 2·8 per cent if the percentage in the 182 to 190 shillings group is assumed to be divided into five equal parts.

The psychologist would also be interested in the time trend of the sale and the situation of the buyer at the sale. Since places of lots in the sale were allotted by ballot, it would be a reasonable assumption that quality was randomly distributed throughout the sale. But what of buyer behaviour? How were price levels affected by buyers arriving late, buyers departing for lunch, and buyers leaving having bought what they wanted?

To provide answers to these questions we divided the sale list into five equal parts of 43 lots each and found the average price per lot, expressed as shillings per head, in each part. Table III gives the results of this.

TABLE III

Average price per lot at various stages of the sale

Period	A	B	C	D	E
Average price (shillings per head) ..	194·6	206·4	199·1	192·0	193·5

The price was below average in the first part of the sale, probably as buyers were still arriving, or holding off until they saw what prices were like, and depressing the price in doing so. In the next part of the sale, probably with the great majority of buyers present, the price rose twelve shillings above that of the previous part and almost ten shillings above the average for the sale. In the third period the average price dropped seven shillings, although it remained above average, and dropped again to give the two lowest average prices in the two final periods of the sale. Table IV gives the results of an analysis of variance to test the significance of these differences. The figure of F found is not quite significant at the five per cent level, but is sufficiently close to suggest that the differences may be real.

TABLE IV

Analysis of differences of prices at various stages of the sale

	Sum of Squares	Degrees of Freedom	Mean Square	F
Between periods ..	5 717	4	1 429	2·2
Within periods ..	137 764	210	656	
	143 481	214		

Besides the possibility of buyers departing satisfied, which would probably account for the low average price in the last two periods of

the sale, there was probably the further effect of buyers departing during the sale for lunch. The sale began at 10 a.m. and lasted until about 3 p.m., so that each fifth part corresponded roughly to one hour of the sale. The lunch period 12 noon to 2 p.m. would extend over the third and fourth periods, and would account to some extent for the drop in the average price after the second period.

Although the evidence just presented relates to the behaviour of demand in a market situation, the data so far presented have been for the interest of the psychologist and the statistician, and the economist may be excused for wondering what has become of the study of the 'approximation to the perfect market'. The first question to be answered for the economist is definitely: 'Where is the demand curve, and how can it possibly be constructed from the set of data that provided Table I?' One answer appears to be that the question being asked is 'How many sheep were sold at or above price y?', and the answer to that is a graph with price on the y-axis and cumulative

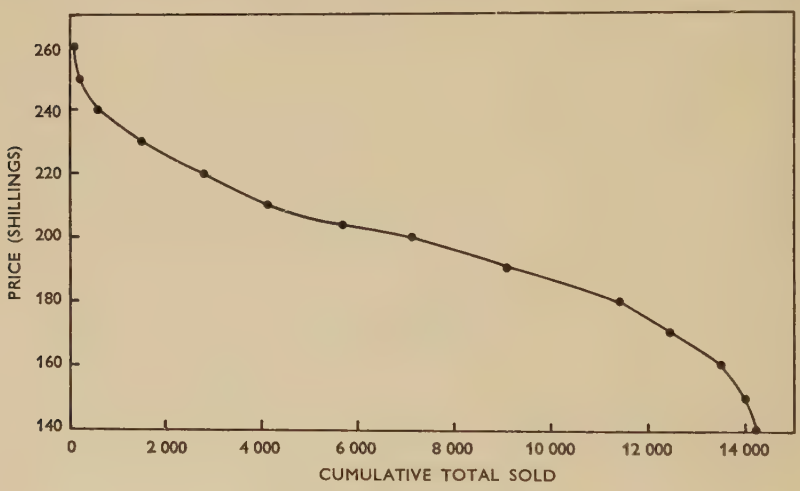


Fig. 1. 'Demand' curve for sheep sold at auction sale.

number sold on the x-axis. This is shown in Fig. 1, which was constructed from the raw data with reference points at every ten shillings. It will be seen from this graph that the demand curve which it represents varied in elasticity from inelastic in the upper price regions to elastic in the middle regions, but below the mean price the elasticity reversed and the demand curve became inelastic again. One reason for this would appear to be that the diagram deals with a finite quantity that has been sold, and the demand curve must therefore cut both axes. The classical demand curve, since it refers to a hypothetical situation, usually cuts neither. More important, our curve is not, strictly speaking, a 'demand curve' but a 'demand and supply curve'. The goods were provided by sellers who were selling of their best, and therefore a restriction of supply probably operated in the lower price

ranges. This might be taken to the point where, rather than sell at a price which he considered much less than just, a seller withdrew his goods from sale. In the sale mentioned here, only two lots listed on the catalogue were not sold, for reasons which we do not know. In passing, we would also remark that the bias in the median towards the higher price levels has the counterpart of a relatively short high-price range over which demand was inelastic. This may have had something to do with general market conditions since the sale studied was part of a 'rising market'. It would be interesting to see the results of similar sales in a 'falling market' so as to test the hypothesis that the median would have a bias towards the lower prices. The percentage of lots withdrawn from sale would also have to be taken into consideration.

However, the economist is unlikely to accept this demand curve as representing an approximation to a perfect market. The approximately normal distribution of sales over the price range given in Table I has already been attributed to the variation of quality in a known fashion, so it cannot also be said to be due to variations in demand and supply. Before conclusions can be drawn about the action of demand and supply, differences of price due to quality must be averaged out. In this case it is a reasonable assumption that quality was randomly distributed throughout the sale, so in fact the basic data for a simplified 'real' demand curve are contained in Table III. If these figures are arranged in the order B, C, A, E, D, and put on a graph with price on the y-axis and cumulative number sold on the x-axis, the result is a demand curve which on the whole is elastic, and most so in the lower price range.

I conclude with a word of warning. Industrial psychologists are cautious beings, perpetually reminding others that one swallow does not make a summer and that it is dangerous to generalise on the basis of only one instance. In keeping with our character, I must therefore point out that my concern here has been merely to investigate and describe in the interests of disseminating knowledge and provoking thought, an isolated instance of an auction sale, and to speculate on possible hypotheses which might be tested in further research into the subject. Such auction sales are held several times a year in many parts of the country, and probably show individual differences according to region. Even the same sale may differ from year to year in the results, according to market conditions, and even according to the weather. But these of course are only further reasons why field investigations of market conditions should not stop here.

I wish to acknowledge with thanks the assistance of John Swan & Sons Ltd., Newtown St. Boswells, in providing the data on which this account is based.

A REGRESSION CONTROL CHART FOR COSTS

EDWIN MANSFIELD and HAROLD H. WEIN

Graduate School of Industrial Administration, Carnegie Institute of Technology, Pittsburgh, Pennsylvania

Although the idea of a regression control chart for costs is not new, published examples of its use are difficult to find. In this article the authors describe the use of a multiple regression equation to establish a cost control chart designed to help the management in the efficient operation of a freight yard.

THERE have been numerous studies where the regression of cost on output has been estimated for manufacturing plants.[1] These studies fall generally into two categories. In the first category, the chief purpose of the work has been to accumulate evidence concerning various cost curves that occupy an important place in economics. In the second category, the work has been done primarily with a view toward the direct usefulness of the regression to industrial managers. This paper falls into the second category.

We give an illustration here of how such a regression may be used to formulate a simple control chart for costs, a topic which should be of interest to many applied statisticians and industrial managers. The problem of controlling performance and costs is important in almost every sector of industry and trade. Our data, and hence our specific results, pertain to the American railroad industry, but the statistical techniques that are used may be applicable to other industries as well.

The plan of the paper is as follows. The first section provides some introductory material concerning the railroad industry and the problem of controlling costs in freight yards. The second and third sections describe the regression control chart and how it is constructed, and the fourth section deals briefly with the performance of the chart. The remaining sections contain a discussion of the differences between this chart and those typically used in quality control, and some concluding remarks.

Freight Yards and Cost Control

Freight yards differ greatly in size, layout, and type, but they have certain physical characteristics and functions in common.

Physical Characteristics. All yards contain sets of tracks. In large yards they are generally of three types: receiving tracks where incoming freight cars are stored, classification tracks where cars are switched, and outbound tracks where cars that were situated on a classification track are stored until a locomotive hauls them away as a train.

Functions. Freight yards switch cars. That is, they sort incoming cars by putting them on the appropriate classification tracks, and in this way they break up incoming trains to form new trains. Most yards also deliver and pick up cars. Engines are assigned to deliver cars to industrial sidings and other yards and to pick them up there. Finally, many yards bill and inspect freight cars and perform such ancillary services as maintenance, repair, and storage.[2]

The importance of freight yards to the railroad is illustrated by the fact that about one-third of its total operating costs may arise in these yards. In view of this, it is clearly important that proper control be maintained over the performance of the yards. However, this problem of maintaining adequate control is made difficult by their number and their distribution over a large area. (For example, in the railroad we studied, about 200 freight yards are scattered along approximately 12 000 miles of track.) It is virtually impossible for any team of managers to have each day a reasonably complete knowledge of what happened at each yard. They must examine selected data concerning the performance of the yards during the day, and from these data they must somehow evaluate a yard's performance. In evaluating performance, one piece of information that is used is the costs incurred in the yard during the day.

At present, the data and techniques used by most railroads do not seem well suited for their purpose.[3] Judging by the opinions of the railroad management with which we worked, their purpose is to detect those days when the costs at a yard are unusually high for the output produced and those series of days when the costs are repeatedly higher than would be expected. Detection of either of these would result in an inquiry concerning the causes of the apparent deterioration of yard performance. In addition, the management is interested in detecting days when costs are unusually low or when they are repeatedly lower than would be expected. In this case there would also be an inquiry, but the intention would be to encourage the responsible factors rather than to remedy them.

In this paper we discuss a control chart based on the regression of cost on output that may be useful for these purposes. Before discussing the chart, it seems worth while to describe the measures of freight-yard output and costs that are used. The two most important services performed at a yard are switching and delivery; and it seems reasonable to use the number of 'cuts' switched and the number of cars delivered during a particular period as a measure of output. A 'cut' is a group of cars that rolls as a unit on to the same classification track; it is often used as a measure of switching output.[4] The number of cars delivered includes both the cars delivered to sidings and other yards and those that are picked up. This output measure is not ideal, one difficulty being that it conceals considerable heterogeneity. For example, two groups of cars may be delivered but one may be hauled a greater distance than the other. Some of this heterogeneity could be eliminated by further

refinements in the output measure, but the extra complexity with regard to data collection and computation might result in a loss of feasibility.

The costs used here include all money costs incurred in the yard except fixed charges, repair costs, maintenance and storage costs, and vacation costs. Only money costs are included; the costs that may be imputed to car delay are not taken into account. Fixed charges are excluded, but some of the included costs are essentially fixed in the very short run.

The Regression Control Chart

The regression control chart contains the residuals from the regression of cost on output, and it uses these residuals to detect days when costs are suspiciously high or low. The statistical model that underlies the chart is as follows: for a particular yard, the cost on the i'th day (C_i) is viewed as a normally distributed random variable. The expected value of C_i is assumed to be a linear function of the number of cuts switched on the i'th day (S_i) and the number of cars delivered on the i'th day (D_i). The C_i are assumed to have a constant residual standard deviation and to be serially independent. In short,

$$C_i = \alpha + \beta_1 S_i + \beta_2 D_i + z_i \qquad \ldots.(1)$$

$$\mathrm{E}(z_i) = 0 \qquad \sigma_{z_i} = \sigma \qquad \mathrm{E}(z_i z_j) = 0 \qquad \ldots.(2)$$

where z_i is distributed normally.

When the railroad management refer to unusually high or low costs, it seems clear that they mean costs that are unusual if the cost-output relationship and the effects of numerous small disturbances remain at their previous, satisfactory levels. That is, they are interested in detecting those C_i that are unusually high or low if the values of α, β_1, β_2, and σ are unchanged. Similarly, when they refer to a sequence of days when costs are higher or lower than would be expected, it seems clear that they mean a run of the C_i that is unlikely if these parameters are unchanged.

If the model is adequate and if the previous values of α, β_1, β_2, and σ are known, it is a simple matter to set up a regression control chart that will aid the management. Each day, $C_i - \alpha - \beta_1 S_i - \beta_2 D_i$ can be plotted on a chart that has two sets of control limits. The outer control limits can be set so that, if α, β_1, β_2, and σ remain the same, the probability that a point lies outside them is small. The inner control limits can be set so that, if the parameters remain the same, the probability that two consecutive points lie outside them (in one direction) is small. When a point lies outside the outer limits or a pair of points lies outside the inner limits, there is evidence that the parameters may have changed and that an inquiry should be made.

Of course, the previous values of α, β_1, β_2, and σ are seldom known when the chart is first set up and put into use. Estimates of these values must be used instead. In setting up the chart we treat these estimates

as if they were the values of the parameters. That is, we ignore the sampling errors that attach to the estimated regression. Moreover, in judging whether the previous performance was generally satisfactory, the management must use the estimates rather than the parameters. The substitution of the estimates for the parameters in formulating the chart results in obvious errors that may be bothersome when the chart is first established. However, as more data are accumulated they become less important, since increases in sample size tend to reduce the sampling errors. (Of course, the sampling errors are irrelevant if the estimates themselves are judged to be satisfactory values of the parameters and if the chart is used to detect differences between the actual values and the estimates.)

Finally, a linear regression seems reasonable on *a priori* grounds, since these yards seldom operate near capacity. The data used below seem to indicate that it is adequate, but no very extensive tests were carried out. Taking account of computational feasibility, it is important, of course, that there should be only a small number of independent variables.

Setting up the Chart

During the summer of 1956 a regression control chart was set up at a freight yard located at Toledo, Ohio. This yard constitutes one of the largest and most important links in the railroad we studied. The chief types of freight that pass through the yard are livestock, perishables, coal, and automobiles. Table I shows the number of cars

TABLE I

Output, cost, and employment, freight yard, Toledo, 7 days, 1956
(*Taken from records of co-operating railroad*)

Item	Fri.	Sat.	Sun.	Mon.	Tues.	Wed.	Thurs.
Number of cuts switched ..	869	792	762	586	669	732	659
Number of cars switched ..	2 534	2 303	2 521	1 849	2 090	2 114	1 979
Number of cars delivered*	1 015	1 003	820	548	877	706	1 038
Number of crews used ..	45	45	40	38	46	46	46
Number of engine hours ..	372	369	329	309	385	381	386
Money costs ($)	7 523	7 464	6 932	6 550	7 606	7 757	7 701

* Includes number of cars picked up.

switched, the number of cuts switched, the number of cars delivered, the number of crews employed, the number of engine hours used, and the costs at the yard, for a sample one-week period.

The first step in setting up the chart was to gather historical data concerning cost and output. Data similar to those in Table I were collected for sixty-one days, and the regression of cost on output was estimated. The resulting regression equation was

$$C_i = 4\,914 + 0{\cdot}42S_i + 2{\cdot}44D_i$$

Standard errors of estimates: 0·5 0·4 ....(3)

The estimated values of the parameters were judged to be satisfactory. The standard errors of these estimates are large, and they limit the initial usefulness of the chart. But as more data became available they became smaller and somewhat less important. There are strong *a priori* grounds for including the number of cuts in the regression, and this variable was included despite the fact that its regression coefficient was less than the standard error.

The second step was to test some of the assumptions underlying the chart. Some of these tests are quite similar to those used in quality control to determine if the process is 'in control'. The statistical model states that the residuals from the true regression are serially independent, normally distributed, and of constant standard deviation. The true regression is unknown, and the residuals from the estimated regression (shown in Fig. 1) must be used in the tests.

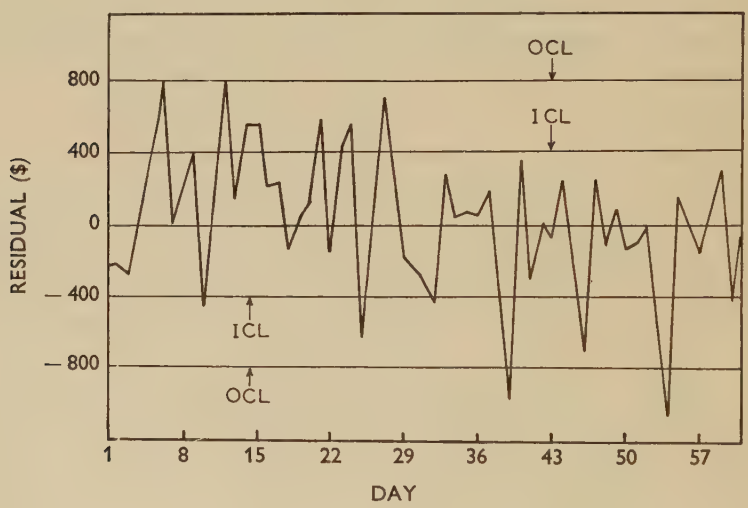


Fig. 1. Residuals from regression of cost on output: freight yard, 61 days, 1956. (Source: records of co-operating railroad.)

Four tests were used to detect serial dependence among the residuals. (1) A conventional runs test was conducted. There are 30 runs above or below the median; the longest run is 6 observations. Both of these results are non-significant at the 0·05 probability level.[5] (2) The Von Neumann ratio, δ^2/S^2, where δ^2 is the mean square successive difference and S^2 is the variance, was computed and it equalled 1·78, a value that is non-significant at the 0·05 level.[6] (3) The residuals were correlated with time, and the correlation coefficient (−0·30) was non-significant at the 0·01 level. (4) An analysis of variance was conducted to determine whether there were significant differences among the mean residuals corresponding to the days of the week. The analysis of variance table may be summarised:

Source	Degrees of Freedom	Mean Square
Between days of the week ..	6	432 336
Within days of the week ..	54	122 436

The mean-square ratio (3·53) is significant (0·01 level). This is because the residuals on Sunday tend to be large and negative.

To test for deviations from normality, the moment ratio ($\sqrt{b_1}$) and the ratio of the mean deviation to the standard deviation (w) were computed. Their values were $\sqrt{b_1} = -0{\cdot}31$, $w = 0{\cdot}77$. Both are non-significant (0·05 level).[7] In addition, a statistic (t, but not Student's t) recommended by Kendall was computed, and it too was non-significant (0·05 level).[8] Of course, these tests, based on 61 observations, are not very powerful. Finally, an analysis of variance was used to test whether the variances of the residuals differed between the two months (the first 31 and the last 30 days). The variances were 147 182 for one month and 132 810 for the other. The difference proved to be non-significant (0·05 level). This was only a crude test of the assumption of constant residual standard deviation, but the results did not contradict the assumption.

Taken together, the results of these tests did not cast any great suspicion on the assumptions underlying the chart. Indeed, the results seemed to be quite compatible with these assumptions. The only piece of evidence that seemed to conflict with them was the very low level of the residuals on Sundays, and we and the railroad staff, after a considerable search, could find no reason for this. We finally decided that it was advisable to include the Sunday residuals when estimating the standard deviation.

The third step was to draw the inner and outer control limits on the chart. The estimated standard deviation, $\hat{\sigma}$, of the residuals was \$410. The outer control limits were set at $\pm 1{\cdot}96\hat{\sigma}$ ($\pm$\$804) and the inner control limits were set at $\pm\hat{\sigma}$ ($\pm$\$410). These limits (designated by ICL and OCL) are included in Fig. 1. If there were no errors in the assumptions and the estimates of the parameters, the probability would be 0·05 that a point would lie outside the outer limits if the parameters remained fixed. Similarly, the probability would be about 0·03 that two consecutive points would lie outside the same inner control limit. Of course, these control limits are only one possibility out of many; the probabilities may be altered to meet the particular situation encountered. Moreover, the use of the inner limits is only one possible way of detecting suspicious runs.

After setting the control limits, an attempt was made to determine if any assignable cause could be found for the days that were 'out of control'. None could be found, and it was assumed that they were due to 'chance'. The number of such days was almost precisely what one would expect on a chance basis.

At this point, then, the parameters of the regression had been estimated, the assumptions of the model had been tested, and an attempt had been made to determine assignable causes for the days that were out of control. Note again that no assignable causes could be found. In particular, the factors cited in the next section to explain the low residuals on Sundays during a subsequent period could not explain them during this period. The next step was to use the control limits to control yard costs. This step is described in the following section.

Performance of the Chart

This section describes the performance of the chart during a six-week period that was several months subsequent to the time when the chart was set up. The results apply to the freight yard described above. On each day during the period data were collected concerning the money costs (C), the number of cuts switched (S), and the number of cars delivered (D) on the previous day. Then $C - \hat{\alpha} - \hat{B}_1 S - \hat{B}_2 D$ was computed and plotted on the chart. The regression was recomputed at the end of each week, $\hat{\alpha}$, $\hat{B}_1$, and $\hat{B}_2$ being estimates based on the data underlying Fig. 1 and data gathered subsequent to the formation of Fig. 1 but before the week when the chart was used. Observations where there were assignable causes were omitted from the latter data. The residuals that were plotted are shown in Fig. 2. The control limits

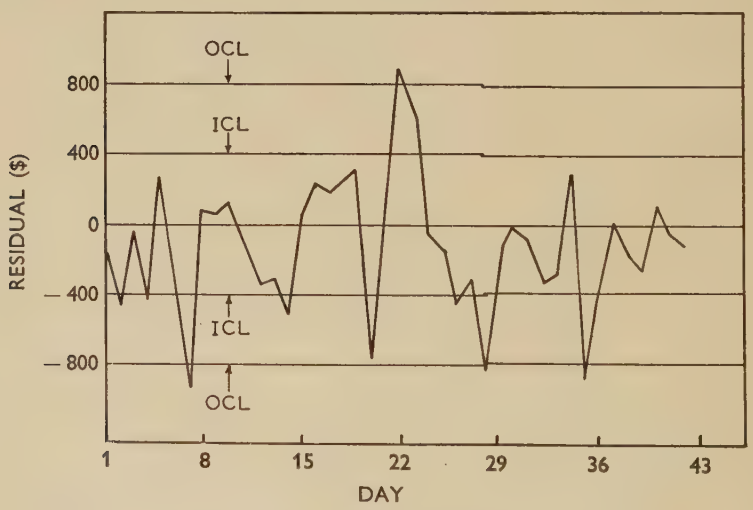


FIG. 2. Regression control chart for costs: freight yard, 42 days, 1956. (Source: records of co-operating railroad.)

differ slightly from one week to the next because the regression and the limits were recomputed each week, but the differences are seldom large enough to be noted.

During the six-week period, four days seemed to be out of control, and in every case there seemed to be an assignable cause. One of these

days was Labor Day. The exceptionally high costs on Labor Day can be attributed to the punitive wage-rates that were paid because it was a holiday. On the other days that were out of control (all of them Sundays) the exceptionally low costs can be attributed in part to the following circumstances: (1) Cars delivered to other yards constituted a large proportion of all cars delivered, and since such cars are relatively easy to deliver, costs were depressed. (2) More efficient methods were used to handle incoming cars. (3) Some work ordinarily performed by the yard was done by another yard; hence costs were reduced somewhat.

The basic data were collected by railroad employees and the points on the chart were computed by officials of the railroad. From their evaluation of the performance of the yard it appeared that the chart provided a reasonably faithful picture of the level of performance. Moreover, the chart stimulated some inquiries that indicated where improvements in yard performance might be made.

To increase further the usefulness of the chart, it appeared that cars delivered to other yards might be separated in the output measure from cars delivered to industrial sidings. In this way an important factor (and in some ways an extraneous one) responsible in part for the low residuals on Sundays would be taken into account. No finality can be claimed for the results, but it appeared that the chart would be useful to the management in controlling yard costs.

Differences from Quality Control Charts

It seems worth while to call attention to (1) certain differences between this chart and those commonly used in industrial quality control and (2) the problem of maintaining continuity in the chart in the face of wage-rate changes. The regression control chart differs from the typical quality control chart in at least four important respects:

First, single observations rather than the means of small samples are plotted on the chart. Single observations must be used, because the management is interested in day-by-day performance. The use of single observations makes it important that the distribution should actually be close to normal; no longer can an appeal be made to the central limit theorems.[9]

Second, a regression plane rather than a mean is used as a measure of central tendency. Because a regression is used, the errors in the chart arising from the use of estimates (rather than parameters) vary from day to day. In the typical quality control chart, limits are set on the basis of estimates of the process mean and standard deviation. The estimates contain sampling errors, but the latter do not vary from day to day. With the regression, however, the sampling errors vary from point to point; they are smallest when S_i and D_i are near their means and largest when S_i and D_i are far from their means.

Third, there is no chart, like that often used in quality control, to detect changes over time in the variability of costs. It is possible that

moving ranges could have been used and that a variability chart could have been based on them. But we decided instead to use F-tests every two weeks or so to test the assumption of constant variability.[10]

Fourth, because the relative costliness of Type I and Type II errors is different from that usually encountered in quality control, the conventional 3σ limits are not used. A Type I error is committed here if one rejects the hypothesis that the parameters are unchanged when this hypothesis is true. A Type II error is committed if this hypothesis is accepted when it is not true. A Type I error is much less costly here than in the typical case in quality control.

Finally, it should be noted that changes in wage rates result in a change in the regression of costs on output. Fortunately, wage-rate changes occur in the railroad industry only at one- or two-year intervals. But it would often be desirable to maintain continuity in the chart despite these changes. The simplest way of doing this is to use constant wage rates in computing costs for the chart despite any actual changes that may occur. Extra clerical work is needed, but experience has indicated that the increase in the clerical load is not very great.

Concluding Remarks

In conclusion, we would add the following remarks concerning the regression control chart and its purpose.

First, the regression of cost on output is not meant to represent the relation between minimum cost and output that economists deal with. Nor is it meant to represent the relation between output and standard cost in some engineering sense. It represents only an average, historical relationship between actual cost and actual output.

Second, the chart may provide information to aid the management in their decision concerning the values of the parameters that are satisfactory. The management may sometimes be unable to specify such values initially or to judge the adequacy of the values that are estimated. In such cases the chart may play a role similar to that often played by the quality control chart. It may be used to explore the present and past characteristics of the 'process'. On the basis of this and other information, certain values may eventually be judged as satisfactory.

Third, the chart does not indicate how to improve performance and it does not indicate why costs on a particular day seem unusually high or low. It only provides a signal. On the basis of the signal, an inquiry must be made to uncover the reasons for the unusual performance and to lead to the necessary corrective action.

Fourth, the idea of a regression control chart for costs is not new. For example, Philip Lyle[11] advanced this idea ten years ago. Surprisingly enough, no application of the idea has appeared (to our knowledge) in the statistical literature. Because the subject is so important in industry the present application may be of interest to many applied statisticians and industrial managers.

Acknowledgements

This research is part of a larger project dealing with the railroad industry. The project was supported by a grant from the Westinghouse Air Brake Corporation to Carnegie Institute of Technology. We should like to acknowledge the valuable assistance of C. Link, R. Nadel, and E. Saunders of the co-operating railroad and the comments of our colleague, J. Walter, on an earlier draft.

REFERENCES

[1] Some of these studies are:

Dean, J. (1941). 'Statistical cost functions of a hosiery mill,' *University of Chicago Studies in Business Administration*, **11**, No. 4, 19.

Idem. (1941). *The Relationship of Cost to Output for a Leather Belt Shop.* Technical Paper No. 2. National Bureau of Economic Research, New York.

Committee on Price Determination (1943). *Cost Behavior and Price Policy.* National Bureau of Economic Research, New York.

Lyle, P. (1946). *Regression Analysis of Production Costs and Factory Operations.* Oliver and Boyd, Edinburgh and London. (Third edition, revised by L. H. C. Tippett, 1957.)

Wein, H. (1947). 'Wages and prices—A case study', *Rev. Econ. Statist.*, **29**, 108.

[2] Beckmann, M., McGuire, C. and Winston, C. (1956). *Studies in the Economics of Transportation.* Yale University Press, New Haven. (A more complete description of yard operations.)

[3] Mansfield, E. and Wein, H. 'Notes on railroad productivity and efficiency measures', *Land Economics* (forthcoming).

Idem. 'A model for the location of a railroad classification yard', *Management Science* (forthcoming).

Idem. 'Linear decision rules and freight yard operations', *J. Industr. Engng* (forthcoming).

Idem. 'A study of decision making within the firm' (mimeographed).

[4] Beckmann, M. *et al. Op. cit.*

[5] Swed, F. S. and Eisenhart, C. (1943). 'Tables for testing randomness of grouping in a sequence of alternatives', *Ann. Math. Statist.*, **14**, 66.

[6] Von Neumann, J. and Hart, B. I. (1942). 'Tabulation of the probabilities for the ratio of the mean square successive difference to the variance', *Ann. Math. Statist.*, **13**, 207.

[7] Geary, R. C. (1935–6). 'The ratio of the mean deviation to the standard deviation as a test of normality', *Biometrika*, **27**, 310; 'Moments of the ratio of the mean deviation to the standard deviation for normal samples', *Biometrika*, **28**, 295.

Pearson, E. S. (1936). 'Note on the probability levels for $\sqrt{b_1}$', *Biometrika*, **28**, 304.

Geary, R. C. and Pearson, E. S. (1938). *Tests of Normality.* Biometrika Office, London.

[8] Kendall, M. G. (1951). *The Advanced Theory of Statistics*, Vol. II, p. 105. Hafner, New York.

[9] Grant, E. L. (1952). *Statistical Quality Control.* McGraw-Hill, New York. (Some disadvantages in using single observations in quality control charts.)

[10] Burr, I. (1953). *Engineering Statistics and Quality Control*, p. 264. McGraw-Hill, New York.

[11] Lyle, P. *Op. cit.* For a discussion of a regression control chart in education see: Sandon, F. (1956). 'A regression control chart for use in personnel selection', *Applied Statistics*, **5**, 20.

INITIAL LETTERS OF SURNAMES

PERCY G. GRAY

Government Social Survey

We believe that the information given in this short note is not generally available, and we hope that some of our readers will find it useful.

It is a fairly common practice of administrators to ask different groups of the population to register or make application for certain things at different times according to the initial letter of the surname. The aim is, of course, to spread the load of work. When one is considering such a scheme it is useful to have an up-to-date distribution of the population by the initial letter of the surname. This note gives such an analysis based on a two-stage sample of parliamentary electors taken from the 1957–58 electoral roll.

The sample consisted of every tenth parliamentary elector taken from a systematic selection of 50 first-stage units in England, 30 in Wales, and 30 in Scotland. The first-stage units consisted of a ward or group of wards, a parish or group of parishes (in England and Wales), or a district council or group thereof (in Scotland). These units have been described more fully elsewhere.[1,2] The results of the analysis are given in the table.

Separate figures are given for England, Wales, and Scotland, and these have been combined, with the appropriate weights, to give estimates for England and Wales, and for Great Britain. It will be seen that in England B is the most common initial letter for surnames, accounting for 11%. In Wales J is the most common, again accounting for about 11% of the surnames. As might be expected, in Scotland the M's are well ahead, accounting for 21% of the surnames.

A closer examination of the first-stage units suggested that there is relatively little variation between the regions of England. Within Wales the rural parts contain an appreciably higher proportion of J's than the towns. The greatest variation occurs in Scotland. For those units representing the crofting counties, the proportion of M's reached nearly 40%.

The results suggest that a division of surnames which appears suitable for Great Britain as a whole may produce peaks of work in Wales and Scotland.

The distribution of electors in England, Wales, and Scotland by the first letter of the surname—1957–58 electoral roll

First Letter of Elector's Surname	England		Wales		Scotland		England* and Wales		Great Britain*	
	%	Cumulative %	%	Cumulative %	%	Cumulative %	%	Cumulative %	%	Cumulative %
A	3·0	3·0	1·7	1·7	2·9	2·9	3·0	3·0	3·0	3·0
B	11·0	14·0	6·5	8·2	8·0	10·9	10·5	13·5	10·5	13·5
C	8·1	22·1	4·2	12·4	8·4	19·3	7·9	21·4	7·9	21·4
D	4·3	26·4	7·4	19·8	5·3	24·6	4·5	25·9	4·5	25·9
E	2·1	28·5	5·3	25·1	1·0	25·6	2·3	28·2	2·2	28·1
F	3·5	32·0	1·8	26·9	3·9	29·5	3·4	31·6	3·4	31·5
G	4·9	36·9	4·2	31·1	5·4	34·9	4·9	36·5	4·9	36·4
H	9·2	46·1	8·2	39·3	5·5	40·4	9·1	45·6	8·8	45·2
I	0·4	46·5	0·2	39·5	0·6	41·0	0·4	46·0	0·4	45·6
J	2·8	49·3	11·4	50·9	1·7	42·7	3·3	49·3	3·2	48·8
K	2·1	51·4	1·2	52·1	2·5	45·2	2·1	51·4	2·1	50·9
L	4·1	55·5	4·5	56·6	3·8	49·0	4·2	55·6	4·1	55·0
M	7·5	63·0	6·7	63·3	21·1	70·1	7·5	63·1	8·9	63·9
N	1·7	64·7	1·1	64·4	1·3	71·4	1·7	64·8	1·6	65·5
O	1·3	66·0	2·1	66·5	0·9	72·3	1·3	66·1	1·3	66·8
P	5·7	71·7	6·6	73·1	3·3	75·6	5·7	71·8	5·5	72·3
Q	0·1	71·8	0·0	73·1	0·2	75·8	0·1	71·9	0·1	72·4
R	4·9	76·7	6·7	79·8	5·6	81·4	5·0	76·9	5·1	77·5
S	9·2	85·9	4·8	84·6	8·5	89·9	8·9	85·8	8·9	86·4
T	4·1	90·0	5·4	90·0	3·3	93·2	4·2	90·0	4·1	90·5
U	0·3	90·3	0·1	90·1	0·1	93·3	0·3	90·3	0·2	90·7
V	0·5	90·8	0·6	90·7	0·2	93·5	0·5	90·8	0·4	91·1
W	8·8	99·6	9·1	99·8	6·0	99·5	8·8	99·6	8·5	99·6
X	0·0	99·6	0·0	99·8	0·0	99·5	0·0	99·6	0·0	99·6
Y	0·4	100·0	0·2	100·0	0·5	100·0	0·4	100·0	0·4	100·0
Z	0·0	100·0	0·0	100·0	0·0	100·0	0·0	100·0	0·0	100·0
	100·0		100·0		100·0		100·0		100·0	
Sample size	22 193		11 322		16 397		—		—	

* Obtained by giving the appropriate weights to the preceding columns.

REFERENCES

GRAY, P. G., PARR, ELIZABETH A. and BLUNDEN, R. M. (1956). *Some Useful Data when sampling the Population of England and Wales.* The Social Survey. (Obtainable from The Social Survey, £5 5s.)

GRAY, P. G. and BLUNDEN, R. M. (1957). *Some Useful Data when sampling the Population of Scotland.* The Social Survey. (Obtainable therefrom, 12s. 6d.)

NOTES AND COMMENTS

We are indebted to Mr E. D. van Rest for the following account of recent developments in the organisation of quality control, which we hope will be of general interest.

The statistician sees quality control as one application of his knowledge of the behaviour of small samples. Before the statistical techniques of sampling, experiment, and decision making can be applied in industry, however, there must be created an opening, a field of exercise as it were, in the manufacturing organisation. It is so with other techniques of value to industry. The task of management is to harness the various techniques available, giving each its necessary pivot of action so that they pull together to a common purpose. It is for this reason that there is a subject deserving the name 'Control of Quality' covering a wider field than the statistical techniques serving it. The statistical control chart is not the whole of quality control any more than the ledger is the whole of accounting.

The manager does, however, need to be convinced of the value to his particular firm of any one of these activities; it would be wrong of him not to use his judgement in this way. He therefore needs some exposition of the nature and purpose of the control organisation which is being recommended as well as of the techniques it will need to use. An excellent form of such an exposition has recently been issued by the British Productivity Council.[1, 2] Their titles describe them sufficiently.

In this field of management, where it is proper that the management interest should predominate, much has happened in Europe in the past four or five years. The rebuilding of industries disorganised by the war has given an opportunity for the introduction of the most recent ideas. The subjects of various conferences arranged over these years by the British Institute of Management, the British Productivity Council, CIOS (International Committee of Scientific Organisations), and the European Productivity Agency have been an indication of this interest.

The policy of the European Productivity Agency, which does not look upon itself as a permanent organisation, has been to encourage the establishment of more permanent organisations to continue to foster the interest in various subjects it succeeds in invoking by its committees, its projects, and its conferences. In accordance with this policy, at the beginning of this year the European Organisation for Quality Control, with headquarters at The Hague, was formed. All European countries have been invited to join this Organisation. In the UK an informal committee already exists, called together by the British Productivity Council, to co-ordinate the activities of various bodies interested in quality control. This committee has sponsored the widening of the scope of the Institute of Engineering Inspection to include quality control engineers in its membership, whether engaged in the traditional engineering industries or not. It is possible that UK membership of the EOQC will be, in line with other countries, made through this national committee. It includes among its members representatives of the

Royal Statistical Society, the Association of Incorporated Statisticians, the Ministry of Supply, the British Standards Institution, the British Institute of Management, the Ministry of Education, the Trades Union Congress, the Institution of Production Engineers, and the Institute of Engineering Inspection.

A bulletin is being issued by the EOQC and distributed through the national committees. Some difficulty may be experienced in backing UK membership with the necessary funds. It is important that the interest of UK industry in this subject and in this form of international co-operation should be expressed; this may be done by asking for copies of the *Bulletin* from the British Productivity Council at the address given in reference 1. The EOQC hopes to sponsor, and later to issue reports on, various projects in the fields of sampling, training, and specification.

Mr C. S. Smith, lately head of the Mathematics Department of Leicester College of Technology and now Vice-principal of the Technical College of Cornwall, is a member of the Board of Management of the new body. To this position he takes the experience of his recent visit to the USA as leader of the European Mission to study methods of training of quality control engineers, and of his subsequent participation in the 1956 EPA Conference in Paris and the planning meetings which followed.

The statistician who enters the quality control field should see his work as a contribution to the solution of a practical problem, the control of a process (the process of manufacture). He must realise that because the problem is a general one—that of variability—no one technique is entitled to the label 'quality control'; he must equally realise that none of his techniques can have fruitful application in industry unless and until a pivot of action has been created for them within the manufacturing organisation.

The fear is often expressed by each side—the management side and the technical side—that in any composite organisation an undue emphasis will be given to the other. Obviously a balanced progress is required, and this the new Organisation hopes to bring about.

REFERENCES

[1] *The Control of Quality* (1957). Action Pamphlet No. 9, The British Productivity Council, 21 Tothill Street, London, S.W.1.

[2] *Applications of Quality Control* (1957). Case Studies No. 3, The British Productivity Council, London.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT GROUP

2nd October 1957. A Sampling Inspection Scheme and its Consequences: R. H. S. Lesser

30th October 1957. Simulation Methods: J. Harling

EDINBURGH GROUP

20th November 1957. The Application of Statistics to Screening Problems in the Pharmaceutical Industry: O. L. Davies

10th December 1957. Printers' Errors: A. Lord

GLASGOW GROUP

23rd October 1957. Statistical Decision Theory: G. A. Barnard

LEICESTER GROUP

3rd October 1957. The Statistician and Personnel Management: B. Ungerson

7th November 1957. Shapes and Sizes of Women's Feet: J. R. Manning

5th December 1957. The Auditor and how he Works: P. N. Wallis

LONDON GROUP

4th October 1957. Profits and Statistics: R. Brech

1st November 1957. A Reliability Study of Sensory Assessments: A. S. C Ehrenberg and J. M. Shewan

MERSEYSIDE GROUP

8th October 1957. Decisions from Samples: J. Ogilvie and J. P. L. Truesdale

12th November 1957. Jump and Wiggle Analysis for Time Series: G. H Jowett

10th December 1957. Time and Methods Study: W. Jones

NORTH-EASTERN GROUP

9th October 1957. The Interpretation of Environmental Hazards: J. H Edwards

Dr Edwards pointed out that in the study of diseases today attention i turning to environmental hazards; he took as his main example the study

of lung cancer. He criticised the use of significance tests on data of this kind and stated that it is more important to estimate the magnitudes of effects and their confidence limits. Among other topics discussed were the estimation of the magnitude of effects in 2×2 tables and the deduction of the frequency distribution of 'dose' in the diseased population from that in the general population, knowing the dose-response relationship.

13th November 1957. Productivity Measurement at the Plant Level: T. E. Easterfield

SHEFFIELD GROUP

24th October 1957. Different Methods of Introduction and Organisation of Quality Control in Manufacture involving Complex Processes: Miss E. A. G. Knowles

28th November 1957. Some Problems of Long-term Research in Industrial Medicine: J. R. Ashford

SOUTH WALES GROUP

14th October 1957. The Epidemiological Approach to Medical Problems: I. T. T. Higgins

Dr Higgins began his lecture with a brief description of the early history of the epidemiological method of studying disease, and gave examples of the way in which the method had been instrumental, during the nineteenth century, in leading to the discovery of the causes of some diseases. He went on to describe some of the work which has been done recently in South Wales. Comparisons between the mortality statistics for diabetics in the Vale of Glamorgan and in the Rhondda Fach were shown, and studies of tuberculosis and pneumoconiosis carried out in Welsh mining valleys were described. Evidence was given to support the thesis that there may be two types of pneumoconiosis.

11th November 1957. Discrimination in Physical Anthropology: M. J. R. Healy

Mr Healy described the problems arising in using the measurements of the teeth of fossil ape-men to assess their relative resemblance to man and to apes, and showed the dangers of using the measurements either one at a time or, without great care, in combination. By means of pairs of canonical variates good discrimination could be obtained between measurements of the teeth of men and apes, and the degree of resemblance of the fossil teeth to either group could be easily and objectively estimated.

10th December 1957. Sequential Inspection and Control of Continuous Processes: W. D. Ewan and K. W. Kemp

The particular requirements of inspection schemes designed to assess the quality of articles manufactured by a continuous process were examined and orthodox control chart procedures were discussed in relation to this problem. A sequential method of controlling quality was described which efficiently utilises the available information in a particular sample and which relates this information to that given by preceding samples. This scheme can be applied to properties measured both as attributes and as variables.

STUDY SECTION

LONDON GROUP

9th October 1957. Current Economic Analysis—New Plans for the NIESR: C. T. Saunders

13th November 1957. A Practical Example of Linear Programming: Miss A. Doig

11th December 1957. Economic Problems of Small and Medium-sized Firms: H. F. Lydall

BRISTOL GROUP

17th October 1957. The Assessment of Therapeutic and Prophylactic Procedures: J. Knowelden

Dr Knowelden mentioned earlier methods of assessing the effectiveness of a treatment, such as the clinical impression and the study of time trends; he then spoke of the constraints imposed on medical experimentation by ethical considerations and discussed the difficulty of obtaining objective measurements. The control procedures employed, including the use of *placebos*, were explained and examples were given from trials of cortisone and the whooping cough and poliomyelitis vaccines.

21st November 1957. Computing Methods for Optimisation Problems arising in Production Planning: C. M. Berners-Lee

Mr Berners-Lee discussed the problem of scheduling a production plan involving combinational problems for which it is impossible to obtain an exact solution in a reasonable time. As examples he took the case of deciding the optimum order in which to perform jobs or make products, the assessment of the penalty for running out of stock, and the assembly line balancing problem. He then showed that methods suitable for digital computers provide good approximate solutions.

MEDICAL SECTION

26th November 1957. Observer Variability in the Use of Respiratory Questionnaires: C. H. Wood and A. S. Fairbairn

BOOK REVIEWS

William T. Morris: *The Rationalization of Industrial Decision Processes* (Engineering Experiment Station Bulletin No. 163). Columbus, USA: Ohio State University, 1957. Pp. iv + 117. $2.

According to the author, this work 'is an attempt to construct an academic foundation upon which the scope and competence of engineering technique may be built and extended'. This target, like the book's title, is ambitious. The problems discussed are interdisciplinary in nature, and involve chiefly philosophy, mathematical statistics, and economics—diverse fields in which, the author tells us, he 'can claim no special competence'.

The book is, however, a reasonable exposition of the decision process at work. It discusses the theory of value for engineering decisions, and mentions indifference curves. It distinguishes between deterministic and probabilistic problems. In the latter case, by assuming the outcome of a course of action to be a random variable, it discusses some of the factors which (rather obviously) influence a decision-maker, through the application of both the Tchebycheff and the Kolmogoroff inequalities. There is a flexible discussion of the decision problem for incomplete information, and a final section deals with the timing of decisions. Sixty-four references to cognate work constitute a useful select bibliography. Typographically, the text is tiresome; it alternates rapidly between open- and close-set print, but the mathematical portions are very clear indeed.

The book sets out to make statements about the nature of engineering decisions, from a variety of criteria and in a number of situations. It does so by building mathematical models of decision-making, distinguishing between descriptive models and those which set out as tools of optimisation. This treatment is, presumably, regarded as 'rigorous'. But it is not self-evident that a development using algebraic symbols is necessarily more rigorous than one using the verbal symbols of our normal vocabulary. To say that there is a buying rule that purchases are only made when the price is less than a is quite clear. There is no intrinsic merit in describing the probability of a purchase under this rule as:

$$\int_0^a f(x)\,dx$$

The book's most ubiquitous concept is that a sequence of decision rules should be applied such that the value of the final state of the system is greater than the value would be if any other sequence of rules were applied. This again is quite clear, even when it is called the 'Axiom of Long-Run Success'. Not a great deal is done for elegance or rigour, unfortunately, with the algebraic version, where a sequence of rules is N, and the situation is d_i:

$$\sum_{i=1}^{n}\{u[N^*(d_i)] - u[N'(d_i)]\} > 0$$

There are, it would seem, two possible reasons for generating a model of this kind. The less powerful is that the algebraic statement may be more succinct and less ambiguous than a verbal one could be. This may justify

the book to some readers: others may find that their usual thought-processes are quite capable of handling the concepts involved without the assistance of mathematical language. The more powerful reason is that the algebraic manipulation of the model might engender valuable and unexpected conclusions. It must be conceded that this *could* happen; perhaps someone will succeed in building on this foundation to make it happen. But it does not happen here. The author sadly notes: 'Nor do any of its conclusions appear to require changes in the theory or technique of present-day engineering work'. It is not perhaps too hard to say that this attempt to discover something new about decision-making by a mathematical analysis has not come off, and to doubt the value of publishing the model that proved so unfruitful.

An example will serve to exemplify the author's general method, and also to illustrate this criticism. Someone has money to invest. He may invest it all in a single opportunity which will either double the capital (with a probability P_1) or lose the capital altogether (with a probability $1-P_1$). Alternatively, he may invest in n independent opportunities on the same conditions. Surely, the considerations affecting this problem are perfectly clear. To talk about them we should probably use words like 'eggs' and 'baskets'. Dr Morris used words like 'the principle of diversification' and 'binomial decisions'. He goes on to invoke the Axiom of Long-Run Success, and to maximise $\mathrm{Eu}(a_i)$, thus:

$$\mathrm{Eu}(a_i) = P_1(2v) + (1 - P_1)0 \qquad = 2vP_1 \text{ (1st case)}$$
$$\mathrm{Eu}(a_i) = 2(v/n)nP_1 + 0(n - nP_1) = 2vP_1 \text{ (2nd case)}$$

'This rather surprising result means that in the long run diversification is of no real benefit to the decision-maker.' A second mathematical argument is given to show that if one considers instead the probability of *total loss*, then the second course is preferable after all. Much would seem to depend on the reader's capacity to be surprised. STAFFORD BEER

J. T. Richardson: *The Reduction and Presentation of Experimental Results*, B.S. 2846 : 1957. London: British Standards Institution. Pp. 43. 10s.

This is in effect a simple little treatise on elementary statistics. It is, as becomes a new *British Standard*, concise and precise. We should not expect to find it exciting.

The author begins by offering some 'rules for the presentation of experimental results' which could usefully be followed. He does, however, beg the question: 'presentation *to whom*?' Only a scientifically competent person will understand the meaning of 'standard deviation', for example, which it is said must *always* be given. The idea that the degree of confidence should somehow be indicated is excellent; but in presenting results to *management* it cannot be assumed that the S.D. can be quoted with impunity. A young worker, then, who followed these rules would please his senior scientist greatly; but he should be warned that his report, set out as recommended, would be rejected as incomprehensible by many managers.

There is a well-written section on the handling of data as frequency distributions, which explains how to group, classify, and graph results. This is followed by an account of the arithmetic mean and S.D., with a section on S.D. estimation from the range. This method, while well known and widely used, is much less reliable in some fields of applied statistics than in others. So, too, is the assumption of normality made throughout the booklet. It

may be that this was not the place to discuss other forms of distribution; and yet in some industrial contexts (for instance, in analysing an experiment made on an accept–reject basis) the *Standard* might mislead anyone who tried to apply it.

There follows a very good elementary discussion of uncertainty in estimation (which introduces the control chart technique), and another of 'rounding' observations and estimates. Both these sections could be studied with great profit by many experimenters.

Considering the modest scope of the book, it accomplishes its task well. One serious omission, for some fields of application, is any discussion of measures of centrality. The mean is simply taken as the most useful measure; but this, like the normality assumption, does reflect a certain bias in the author's notion of what is usually encountered. The median and mode should surely have been explained in a *Standard* bearing this title, and their definitions included in the Glossary.

This brief Glossary contains only ten terms. A final doubt concerns the inclusion of 'standard error' as a term 'synonymous with standard deviation, but generally used in referring to the standard deviation of the distribution of values of a statistical measure (e.g. the mean) rather than to that of individual observations'. It is perhaps my personal bias that finds this infelicitous; the term is widely used. But in applied statistics we usually emphasise that variability is a natural phenomenon; it is in fact what statistics is *about*. One of our aims is to convince people that uncertainty is *not* error, but a fact of nature with which we can cope. The term is one we could do without.

STAFFORD BEER

J. Neyman (Editor): *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. V. Econometrics, Industrial Research, and Psychometry.* Berkeley: University of California Press (Agents: Cambridge University Press), 1956. Pp. ix + 184. 43s. 6d.

The three Berkeley Symposia, one in 1945–6, the second in 1950, and the third in December 1954, have had the great value of bringing together statistical problems from very different fields, problems which prove to have a great deal in common. The papers are in effect short monographs on these common problems.

The volume under review is one of five and contains ten papers in all. Although only the three in the section on Contributions to Industrial Research will be dealt with here, some of the others, for example Statistical Inference in Factor Analysis, and Problems of Value Measurement for a Theory of Induction and Decisions, will also be of interest to readers of *Applied Statistics*.

The first of the three reviews the several continuous sampling plans which have been published, plans which alternate total with fractional inspection from an output in order of production, and suggests the direction further work should take. It appears that in the early schemes, and for these Dodge and the Bell Telephone Laboratories must be given the credit, choice of the fraction f to be inspected and the number i of good articles required before changing to fractional inspection were made on the basis of a desired average outgoing quality for a fixed proportion defective; that is, the process was supposed under control. Here it is pointed out that the upper limit to this AOQ, the AOQL, is easily calculable but may in practice be so much higher than the AOQ as not to be of practical interest. Other schemes smooth out

the transitions from fractional to total inspection by requiring more than one defective article before making the change. Yet others make it possible to prescribe a realistic AOQL.

Two points are strongly made; first, that all these schemes should be studied against the background of a realistic production process and, second, that there should always be some sort of economic penalty, such as the additional cost of total inspection over the fractional, to give the incentive for good quality. Several types of plan and their properties are described and suggestions are made for the direction further research should take.

Industrial experimentation, especially in the process industries, is characterised by a great deal of variation attributed intuitively to a large number of sources. The factorial experiment is thus appropriate, at least in the exploratory stages of seeking optimal conditions. In spite of the comparative rapidity with which results can usually be obtained in industrial experiments the numbers of tests required, multiplied as they often are by a number of properties to be measured, become prohibitive. Fractional replication can be resorted to, with its accompanying confounding of the less important effects. The author of the second paper restates the usual notations and sets out in a condensed form the simple rules for writing down the plans for fractional replicates and recognising the aliases.

A suggestion for the quick estimation of errors and the recognition of the important effects is to plot the magnitudes of the various contrasts on arithmetic probability paper. Suggestions are made for dealing with the absence of any combinations which prove physically impossible of achievement. Duplication of some cells only of the factorial design at an early stage might be used to obtain a preliminary estimate of error. Certain sets of fractional replicates that have been found useful in sequential experimentation are discussed.

The third paper, entitled 'Some sequential procedures for selecting the best exponential population', deals with the problem implicit in life-testing of deciding with as few tests or as short an elapsed time as possible which is the best of the several populations from which the samples are taken for test. The means of reducing the time required for life tests (excluding that of increasing the severity of the tests) are to increase the number tested, to replace failures by new units or to use an appropriate sequential test. Such tests are given and their efficiencies compared.

The value of such periodic reviews of subjects in which advances are taking place cannot be over-estimated, but for the benefit of the wider circle unable to attend at Berkeley it could be wished that the interval before publication were shorter.

E. D. VAN REST

U. Grenander and M. Rosenblatt: *Statistical Analysis of Stationary Time Series*. New York: John Wiley & Sons, 1957. (Stockholm, Almqvist and Wiksell; London, Chapman and Hall.) Pp. 300. 88s.

This book is an important contribution to the literature on time series analysis. It is, however, much more theoretical in character than one might expect from its title, and although there is some discussion of applications to statistical problems in physics and communication engineering, most of the book consists of an exposition of the mathematical theory underlying the techniques which are presented. It is not, therefore, a book for the statistician who is primarily interested in the methods available for analysing

ime series so as to be able to apply them successfully in some particular eld, and seems likely to be useful chiefly to those engaged in research with he object of discovering new methods or placing existing ones on a more atisfactory footing. A fairly advanced knowledge of mathematical statistics s essential for the reader (the authors mention H. Cramér's *Methods of Mathematical Statistics* as an indication of the level required), and even with his he will probably find it a difficult though stimulating book, containing nany arguments that can be fully understood only after careful study.

The main feature of the book is its emphasis on spectral analysis. This omprises techniques for making inferences directly about the spectrum or pectral distribution function $F(\omega)$ of the stationary stochastic process $\{x(t)\}$ vhich serves as a model for the time series; these have the advantage of not equiring a detailed knowledge of the form of F and so may be considered non-parametric'. Knowledge of the spectrum is equivalent to knowledge f the correlational structure of the process: for example when the possible alues of t are integers, as in the case of observations separated by equal ime intervals, the covariances $r(s) = \operatorname{cov}\{x(t), x(t+s)\}$ can be expressed as ourier-Stieltjes integrals

$$\int_{-\pi}^{\pi} e^{is\omega} dF(\omega)$$

ut correlation analysis is less satisfactory except when the structure is ompletely specified apart from the values of a finite number of parameters. loreover the solution of many problems is most easily expressed in terms of F.

The importance of the spectrum is made clear in Chapter 1, which deals ith the basic probability theory required for handling stationary processes, nd Chapter 2, which is concerned mainly with problems of linear prediction, nterpolation, and filtering (using an approach similar to that of Doob in is book *Stochastic Processes*). Chapter 3 surveys methods of analysis based on arametric models, for example autoregressive or moving average schemes f some specified order; these are considered less useful, at any rate for roblems in the physical sciences, than the non-parametric spectral analysis echniques.

Chapter 4 is devoted to a detailed discussion of the problem of estimating (ω) or its derivative $F'(\omega)$, the spectral density. Suitable estimators of two ifferent types are obtained, and their properties are investigated on the ssumption that the sample size is fairly large. Some applications, including he determination of spectra of random noise, turbulence, and storm-enerated ocean waves, are considered in Chapter 5. In Chapter 6 the large-ample distribution theory for estimators of the first type (which are linear unctions of the sample autocovariances) is developed, and it is shown how, y means of this, asymptotic confidence bands for $F(\omega)$ and $F'(\omega)$ can be onstructed.

It may be necessary to include a trend term, $m(t)$ say, in the model, so iat now $x(t) = m(t) + y(t)$, where $\{y(t)\}$ is a stationary process. Chapter 7 eals with the estimation of $m(t)$ when this is of the form $\Sigma c_r \phi_r(t)$ where ie ϕ_r are known functions of t. It is shown that under certain conditions, hich are satisfied, for example, when the ϕ_r are polynomials or trigono-ietric polynomials, the least-squares estimates of the c_r, computed on the ssumption that the $y(t)$ are uncorrelated, are asymptotically as good as the inimum-variance unbiased estimates.

The final chapter contains a rather sketchy treatment of some miscellaneous problems, most of which call for further work (among them may be mentioned the effect of substituting an estimate of the spectrum in the formula for the optimum linear predictor). This is followed by a set of exercises for the reader; many of these provide extensions of the theory given in the text.

I feel that the presentation of many of the important results is unnecessarily formal and could have been made clearer without appreciable loss of rigour. There are also places in which the development of the mathematical theory might with advantage have been abridged—for example the earlier part of Chapter 6, where much of the analysis is reproduced from a paper by the authors in the *Annals of Mathematical Statistics*—so as to leave more space for the discussion of applications. Nevertheless, since the book is in effect a research monograph based on the authors' own work during the last few years, these are not to be regarded as serious criticisms. A. M. WALKER

W. J. Reichmann: *The Fascination of Numbers*. London, Methuen & Co., 1957. Pp. 176. 15s

This book is not intended to be used as a textbook but is designed to be used by the general reader as a source of recreation and instruction. Consequently little more than an elementary knowledge of mathematics is assumed, and where proofs are given, the notation is kept as simple as possible. Many of the usual topics of number theory are included. Among them are scales of notation, square numbers, triangular numbers, hexagonal numbers, etc., digital roots, prime numbers, divisibility, perfect numbers, recurring decimals, and so on. In addition there are chapters on various methods of multiplication, magic squares, fallacies, and other things.

In some places I found the exposition rather loose and therefore not entirely satisfactory. For example, a method was given for the calculation of a special class of Pythagorean numbers, and then a formula for generating all Pythagorean numbers. I think it would have been logically more satisfying to state the second formula and to derive the first as a special case. One or two errors came to my notice: some inconsistency of symbolism on p. 13 and a 'Sin' omitted from equation (iii) on p. 84.

It is perhaps significant that reference to statistics is made in the chapter on fallacies. In the words of the author: 'all statistics have their inherent faults and must be treated severely on their merits'. With this somewhat enigmatic statement I leave this review; the book may appeal to those who like to dabble in numbers, for use on long winter evenings. D. G. BEECH

R. A. Buckingham: *Numerical Methods*. London, Sir Isaac Pitman & Sons Ltd., 1957. Pp. xii + 597. 70s.

Among books on its subject, this one is outstanding for its comprehensiveness. The various effective procedures for each type of problem are described in turn, and their merits are compared. There is a wealth of references including full references to the latest literature. Also the numerous tabular appendices of coefficients and formulae will be found very useful. The style is well condensed yet clear and easy to follow, and, wherever desirable, the methods are illustrated by worked numerical examples. I must, however, protest about the print; though clear, this is most unpleasing to the eye.

In scope, the book covers in detail almost every aspect of the science of hand-computation. (The only important exceptions are Fourier analysis and

armonic analysis, for which, though they are outlined briefly, the reader is eferred to other works, which are fully listed.) It gathers together much ecent material hitherto scattered in journals.

Chapter 10 (on the fitting of data by least squares) is the only chapter irectly connected with statistics. Nevertheless, since the applied statistician constantly faced by problems of computation, he will find much of value o him throughout the book, and especially in the chapters on solution of near equations, inversion of matrices, etc.

Chapter 1, entitled 'An Introduction to Computation', deals with errors, ables, and mechanical aids; chapter 2 with differences and sums (forward, ackward, and central), factorial polynomials, and factorial moments. The opics covered in the fourteen succeeding chapters may be broadly grouped s follows:

Interpolation and numerical differentiation and integration (for functions f one and two variables); summation formulae (Euler-Maclaurin, etc.); ubtabulation. Fitting of data by least squares. Numerical solution of (non-near) algebraic and transcendental equations. Numerical solution of rdinary differential equations, partial differential equations, and integral quations. Simultaneous linear equations; inversion of matrices; evaluation f latent roots and latent vectors; expansion of λ-determinants. Rayleigh's rinciple. Relaxation methods.

The familiar material on all these topics is included, together with the ore recent (or less-known) developments in them. F. T. M. SMITH

rome C. R. Li: *Introduction to Statistical Inference*. Ann Arbor, Michigan: Edwards Brothers Inc., 1957. Pp. xiii + 553. $7.50.

This book gives a good explanation of the foundations of statistics. It is esigned primarily for students whose mathematical knowledge is limited, nd all but the simplest mathematical proofs are replaced by practical erification, using random sampling from a normal distribution.

After a brief introduction to the average, the standard deviation, and the ormal distribution, the distribution of the average and special cases of the F, and χ^2 distributions are generated by computation from the results of sampling experiment. The application of these distributions to examine omparisons of averages and of variances is explained simultaneously with eir development, and their inter-relationship is given considerable emphasis. he distribution of the regression coefficient b is also determined by calcula-on from the sampling experiment. An explanation is given of the concept nd use of confidence limits for estimated values.

The analysis of randomised one-factor experiments, randomised blocks, nd simple factorial experiments by analysis of variance techniques is out-ned and followed by a description of analysis of covariance techniques. In ch case the procedures are described by examining a simple numerical ample in detail. D. B. Duncan's multiple range test for ascertaining which k sample averages are different when shown to differ by the analysis of riance is the most novel topic in these particular pages. Marked emphasis placed on the use of individual degrees of freedom to form comparisons, ithout sufficient warning that the comparisons should be chosen from tech-logical experience and certainly not based on a knowledge of the results.

The binomial distribution is treated by the normal approximation, leading the χ^2 test for comparison of averages. This is extended to include testing

of goodness of fit and independence. The angular, square root, logarithmi and normal score transformations for use with analysis of variance procedur are described and their application explained. The final chapter is devote to some examples of methods of analysis which are not based on assumptio of normality.

The book is well provided with worked examples, and at the end of eac appropriate chapter there are a good number of examples for exercisi knowledge newly gained. For almost all these exercises no answer is provide

It is not an easy book to read, probably owing to the author's style, nor it free from mistakes. I found eight, of which the most serious was on pag 290, where eight lines have been transposed from the bottom to almost th top of the page. Nevertheless, it is a book which lecturers in statistics shou examine carefully.

Being based on lectures to students of engineering, agriculture, and th pure sciences, it is intended to provide a general introduction to principle which may be followed by a more detailed study of statistical topics of speci interest. Consequently, at 7.50 dollars, it is hardly a textbook one ca recommend students to buy, but must remain a useful source-book for th lecturer who teaches by practical verification. C. J. ANSC

PUBLICATIONS RECEIVED

Marvin Frankel: *British and American Manufacturing Productivity*. University of Illinois, 195 Pp. 130 (paper covers). $1.50.

Productivity in a number of manufacturing industries in 1947–48 compared for Great Britain and the United States, and the factors affectin productivity are discussed under the headings: national resources, plant an equipment, factory size, size of markets, and economic expansion.

Selected British Technical Journals. London: Federation of British Industries, 1957. Pp. 15. 2

This booklet gives a list of selected journals with brief notes on each. It intended to enable engineers outside the UK, particularly FBI Scholars an others trained in Britain, to keep abreast of the latest British engineerin developments by subscribing to the journals best suited to their requirement

Michael B. Jawetz: *Farm Size, Farming Intensity and the Input–Output Relationships of so Welsh and West of England Dairy Farms—A Study in the Application of Marginal and so More Advanced Economic Analysis*. University College of Wales, Aberystwyth, Departme of Agricultural Economics, 1957. Pp. 84. 7s. 6d.

C. A. Moser: *The Measurement of Levels of Living with Special Reference to Jamaica*. Colon Office: Colonial Research Studies No. 24. London: HMSO, 1957. Pp. 196. 13s. 6d.

Bertil Hallgren: *Enuresis—A Clinical and Genetic Study*. Copenhagen: Ejnar Munksgaar 1957. Pp. xii + 159.

Tore Dalenius: *Sampling in Sweden—Contributions to the Methods and Theories of Samp Survey Practice*. Stockholm: Almqvist and Wiksell, 1957. Pp. 247. 20 Swedish crowns.

Carl Henry Alström and Olof Olson: *Heredo—Retinopathia Congenitalis, Monohybrida Recessi Autosomalis*. Lund: Berlingska Boktryckeriet, 1957. Pp. 177 + pedigree charts.

Effortless figuring

With a Monroe 8N-213 you get all the advantages of a fully automatic adding-calculator. Drudgery disappears from routine figure work and even the most complex calculations become merely a matter of feeding figures and pressing easy-to-select keys. Exclusive Monro-Matic features mean that there is less to remember, less to do, less possibility of human error. Here is a machine that cuts costs and saves time in every department where figure work counts.

Monroe Calculators range from simple hand-operated models to the world's most advanced machines. With Monroe you can bank on the best advice on choosing and using your machine and the most efficient and reliable after-sales service.

MONROE

For further particulars, please write or telephone:
MONROE CALCULATING MACHINE COMPANY LIMITED.
BUSH HOUSE LONDON W.C.2. COVent Garden 0211

Titles from an important statistical list

SAMPLING METHODS FOR CENSUSES AND SURVEYS FRANK YATES, F.R.S. 2nd Edn. **42s.**

STATISTICAL METHOD IN BIOLOGICAL ASSAY D. J. FINNEY, F.R.S. **78s.**

THE DESIGN AND ANALYSIS OF EXPERIMENT M. H. QUENOUILLE **42s.**

RANK CORRELATION METHODS M. G. KENDALL 2nd Edn. **40s.**

EXERCISES IN THEORETICAL STATISTICS M. G. KENDALL **24s.**

PROBABILITY AND THE WEIGHING OF EVIDENCE I. J. GOOD **21s.**

BIOMATHEMATICS C. A. B. SMITH 3rd Edn. **80s.**

"GRIFFIN'S STATISTICAL MONOGRAPHS . . . . ."

No. 1 **ANALYSIS OF MULTIPLE TIME-SERIES** M. H. QUENOUILLE **24s.**

No. 2 **A COURSE IN MULTIVARIATE ANALYSIS** M. G. KENDALL **22s.**

CHARLES GRIFFIN & COMPANY LTD

42 DRURY LANE *Established 1820* **LONDON WC2**

BINDING

We shall be glad to undertake the binding of separate parts of *Applied Statistics* into complete volumes at a cost of 20s. 3d. per volume.

Binding cases can be supplied separately at 5s. 6d. each.

Please send enquiries to Binding Department.

OLIVER AND BOYD

Tweeddale Court, Edinburgh, 1

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part 20s.

Annual Subscriptio **Post Free £4 2s.**

Principal Contents of Volume 120 Part 4 1957

Analysis of Geophysical Time Series (with Discussion)

An Analysis of the Slower Temperature Variations at Kev Observatory by Means of Mutually Exclusive Band Pas Filters .. .. *J. M. Craddoc*

Notes on the Specification of Atmospheric Turbulenc *H. Charnoc*

Some Applications of Time Series Analysis to Atmospheri Turbulence and Oceanography *S. Rushton and J. Neuman*

The Relative Numbers and Abilities of some Ten-year-ol Twins .. .. *Frank Sando*

Alternative Presentations of Social Accounting *Harry Burto*

The General Form of the Adding-up Criterion: A Rejoinder b D. G. Champernowne. A Reply .. .. *J. L. Nicholso*

Annual Report of the Council

Proceedings of the One Hundred and Twenty-third Annua General Meeting

Reviews of Books, Statistical and Current Notes, Additions t Library, Serial Publications currently received, Periodica Returns

ROYAL STATISTICAL SOCIETY

21 Bentinck Street - London, W.1

Telephone: Welbeck 7638

Journal of the Royal Statistical Society

SERIES B—METHODOLOGICAL

Price Each Part 30s.

Annual Subscription Post Free 61s.

Contents of Volume 19 No. 2 1957

Problems in the Probability Theory of Storage Systems *J. Gani*

Some Problems in the Theory of Dams *D. G. Kendall* (With Discussion)

Routine Analysis of Replicated Experiments on an Electronic Computer (with Discussion) *F. Yates, M. J. R. Healy and S. Lipton*

Minimax Procedure for choosing between Two Populations using Sequential Sampling *R. J. Maurice*

Sufficient Statistics, Similar Regions and Distribution-free Tests *G. S. Watson*

On the Use of the Normal Approximation in the Treatment of Stochastic Processes *P. Whittle*

The Variance of the Mean of a Stationary Process *E. J. Hannan*

Methods of Construction and Analysis of Serially Balanced Sequences *M. R. Sampford*

The Modified Latin Square *B. Rojas and R. F. White*

Analysis of Variance as an Alternative to Factor Analysis *M. A. Creasy*

Some Further Results in the Non-Equilibrium Theory of a Simple Queue *N. T. J. Bailey*

Normal Approximation to Machine Interference with Many Repair Men *P. Naor*

Stationary Distributions of the Negative Experimental Type for the Infinite Dam *J. Gani and N. U. Prabhu*

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1

Telephone: Welbeck 7638

Scottish Journal of Political Economy

The Journal of the Scottish Economic Society

Contents of Volume V, No. 1
February 1958

Pluralism, Politics and Sociology .. .. *J. A. Mack*
Economic Schizophrenia .. .. *A. K. Cairncross*
The Establishment of the Scottish Steel Industry .. *I. F. Gibson*
Statistical Production Functions and Economic Development *C. E. V. Leser*
The Practical Uses of Input-Output Analysis .. .. *I. G. Stewart*
Reviews:
The Economic Writings of Francis Horner .. .. *A. T. Peacock*
Christianity and Economic Problems .. .. .. *J. F. Sleeman*
Current Topics
Recent Economic Trends:
The Scottish Banks' Interest Rates .. .. *M. Gaskin*

The Scottish Journal of Political Economy is published three times per year, and is free to members of the Scottish Economic Society. The annual subscription to non-members is £1 10s. (U.S.A. and Canada $5). Single copies 10s. 6d. (U.S.A. and Canada $1.50) post free. Orders should be sent to Oliver and Boyd Ltd., Tweeddale Court, High Street, Edinburgh, 1.

URGENTLY REQUIRED

APPLIED STATISTICS

The following back numbers of the journal are out of print. Will readers who have copies which they do not now require kindly advise the publishers.

1952 Volume I Parts 1, 2 and 3
1953 Volume II Part 1
1957 Volume VI Part 1

OLIVER AND BOYD
Tweeddale Court, Edinburgh, 1

EDITORIAL

During the first six years of its publication *Applied Statistics* has contained more than one hundred articles, totalling over one thousand pages. These have accounted for about eighty-five per cent of the text space, the remaining fifteen per cent consisting of reports, book reviews, and other miscellany. This seems a suitable opportunity to comment on the contents from an editorial point of view. Needless to say, we shall always be interested to hear what our readers think.

Articles have ranged in size from two to twenty-one pages, the average being between nine and ten pages. It has been our deliberate policy to encourage articles of medium or short size as being most likely to maintain the right 'atmosphere'. However, long articles have their occasional rightful place, especially as surveys of wide fields; in any event, the chief criterion is comprehensibility.

The subjects dealt with have varied over a wide range; to give some idea of the extent to which various interests have been catered for we have divided the articles into classes. The subdivision is based upon the main appeal of the article and is only rough; it is bound, of course, to be to some extent a matter of personal opinion, since some articles may appeal to more than one class of reader. However, there was usually little difficulty in deciding what was the main topic of discussion. The results, expressed as percentages both of the number of articles and of the number of pages, are as follows:

	Number of Articles (%)	*Number of Pages* (%)
General interest	22	21
Industrial: control of quality ..	12	11
Industrial: design and analysis of experiments	10	10
Industrial: other topics	14	16
Economics and econometrics ..	19	19
Sociology and education	5	6
Sampling and survey work ..	11	10
Medical	6	7

No comment is needed, except to say that the interests of all the various Sections of the Royal Statistical Society are represented. We certainly do not wish to imply that the importance of a subject is to be judged by the amount of our space that it occupies!

We have also subdivided the articles in a different way: into those describing particular applications of statistics, expository articles and surveys, and those describing computational methods and nomograms. The results are as follows:

	Number of Articles (%)	*Number of Pages* (%)
Particular applications	71	67
Expository and survey	21	26
Computations and nomograms ..	8	7

We consider it fitting that the first class should preponderate, since in most cases these articles describe either new techniques or novel applications of old ones; in any case they are specific instances where statistical methods have been usefully applied. But the other types of article also have important parts to play.

So much for the articles that have been published; it may not be out of place to indicate, for the benefit of potential authors, two types of paper that are not suitable. The first type is too academic; it is nothing more than a mathematical exposition of a new and rather complicated technique, which has not been applied. Moreover, no indication is given of possible fields of application. The second type is a simple example of the application of a well-known technique, such as analysis of variance, with no complications, described at great length. It is difficult to justify the publication of such an article, yet the same simple example, clearly and briefly described, might well make a very acceptable contribution.

At the same time, it must be stressed that we welcome articles for publication. We feel that so much statistical work is being done at present that there must be many novel applications of statistical techniques which would help other statisticians if they were brought to their notice in a simple article. The techniques themselves may be new, simple adaptations of old ones, or merely old techniques applied in an apparently unpromising field. Again, for those with a flair for exposition, some techniques may deserve a better treatment than they have received in textbooks. Any success which *Applied Statistics* may achieve depends on an ample supply of material.

Turning now to the miscellaneous material, we believe that it performs a useful function, among other things in helping to preserve the rather informal character of the journal. Two of the features—Letters to the Editor and Questions and Answers—have been sporadic in their appearance, but this is not a matter of policy. We welcome letters giving readers' views on the journal, even if they contain brickbats; comments on, or criticisms of the articles, either on matters of opinion or matters of fact, are also welcome. We do not claim to be infallible. Where publication of letters fulfils a useful purpose, it is our policy to

publish if space is available. It is a matter for regret that more Questions and Answers have not appeared, but again we depend on our readers; send us the questions and we will endeavour to provide the answers.

Finally, it will not have escaped our readers' notice that the amount of detail in the reports of sectional activities varies greatly; most meetings have recently been reported by title only, whereas a few have been given summaries. The main reason for this is that we cannot report details that we do not receive. Although the available space is limited and it may be that some meetings cannot be adequately summarised in a few lines, we should be willing to devote more space to sectional activities so that more meetings could be reported in summary; this was originally intended to be one of the functions of the journal.

There remain a number of items which are most conveniently dealt with by putting them together under the general heading 'Notes and Comments'. Section conferences have been reported in this way, and so have meetings and conferences with some statistical interest arranged by other societies. This is also a useful place to draw attention to forthcoming events, the publication of new journals, the formation of new societies, and other news items. Clearly the section is greatly dependent on the publicity departments of the organisations whose activities are mentioned, but here, again, readers can help by sending us items which may merit a note or a comment.

THE ECONOMIST AND HIS STATISTICS

ERIC SHANKLEMAN

The analysis of published economic data is a function of the growing body of business economists and statisticians. The adequacy of the data for their purposes, some of the pitfalls, and the way ahead for future developments are here discussed in the context of a review of a recent book on British Economic Statistics.

THERE are very many more economists employed *qua* economists today than pre-war: many more in government, many more in public enterprise, which has, of course, been expanded considerably in post-war years, and many more in private industry.

Not all branches of industry or trade have availed themselves of the abilities offered by the economist or, for that matter, his counterpart, the business statistician, equally readily. And it is not always easy to establish a pattern which explains why some industries have, and others have not, added economists to their staff to undertake economics duties. To some extent the degree to which the firm endeavours to programme ahead may determine the need to employ an economist for forecasting. On the other hand, a member of the Board or an executive may prove to be, or may believe himself to be, sufficiently expert in the art of economics to rely on his own judgement, and not to need a specialist on his staff. Said one business economist, due to retire, of the Chairman of his Board, 'He'll not need a replacement in my job; he's too good an economist himself.' The implied economics of this statement are, nevertheless, not beyond contention. These then are some reasons why or why not particular firms employ specialist economists or business statisticians. Another important reason is that the growth of the importance of government departments and economic institutions like Monopolies Commissions, General Agreement on Tariffs and Trade, and now the proposed Free Trade Area, means that in most firms somebody is required to keep under review their operations and their impact on the Company. In the light of these functions there are some somewhat mysterious gaps in the ranks of business economists; thus an attempt to find an economist from within the industry to address the Royal Statistical Society's Conference on Business and Economic Statistics in June 1958 on the outlook for the Machine Tools Industry proved very difficult.

This is not the place to attempt to discuss in detail what the economist or business statistician does in his job. The aim of this article is far less ambitious; it is to look at the material with which he works, and to assess from this review the extent to which he is able to make a

positive contribution to the solution of the firm's or government's problems. The opportunity to carry out such a review occurred with the publication of Professor Devons's book: *An Introduction to British Economic Statistics*[1] and also with Sir Harry Campion's Presidential address to the Royal Statistical Society on 'Recent Developments in Economic Statistics.' But to do this, some idea of the economist's function is clearly necessary—for the material he needs obviously depends on what he is expected to do with it.

These duties can be stated relatively simply. They are to examine the operational economics and the policy of the firm, and to keep under analytical review the impact of external conditions on the firm concerned. Under the first head may be put all the work on internal data from material procurement and production control to sales and demand analysis, calling for varying degrees of statistical sophistication. Included, too, are the assumptions behind price policy, expansion plans, and so forth. But it is the latter function, that of examining the impact of the economic environment on the firm, with which we are mainly concerned in this article. The operative phrase is 'Under analytical review'; and here it is by no means irrelevant to say something about the role of statistics in business.

It is not infrequently emphasised that figures are required 'not for their intrinsic interest, but only to provide a basis for decision. The figures can only provide facts on which judgement is exercised. They cannot replace judgement'.[2] Now, of course, this is true beyond any shred of doubt. But it is probably equally true that the soundness of judgement is largely determined by the experience of the judge; sound judgement, as distinct from the brilliant intuitive hunch, is probably based on sound analysis of past experience—sorting out the relevant aspects of that experience to be applied in the current situation and the present problem. Functionally the task is one of selection and inference. A fascinating study, but one impossible to undertake, would be to score the successful as against the unsuccessful judgements made by the businessman—and equally difficult would be the task of weighting those judgements in terms of their importance to the enterprise. But the challenge to the business economist or the statistician is whether his examination of past experience, of past behaviour, given suitable problems, can lead with efficiency to sound decisions and judgements—efficiency in the sense of the quantity of information required to reach judgements. Complementarily, it is his job to state as reliably as possible, on the basis of this analysis of experience, the probable results of applying alternative policies.

Statistical Tools and Raw Materials

The tools required for this work are technical skill, the raw materials are statistical data; one without the other is clearly valueless. It is to the data that we must turn. Examined from the point of view of the professional business economist or statistician, the aspects with which

we are mainly concerned are the scope of these data, in other words the fields covered (or not covered); the precise aspects of the field tha the statistics cover; and by no means of least importance, the extent t which the figures measure what they are intended to measure.

First as to scope: the development of statistical sources since the wa has been very considerable. Predominantly these have been due t official developments, stemming largely in the United Kingdom an elsewhere from the adoption by the government of a full employmen policy. The proliferation of international agencies, too, has created wealth of valuable data often in a form which eases the business stati stician's task; thus, these volumes are frequently labour-saving in th sense of avoiding the need to consult national statistical compendia Many of the new data available post-war, however, have come fron non-official sources; trade associations merit particular note. But much too, has come as a by-product of the increased number of marke research surveys carried out by various non-government bodies o the characteristics of markets for particular products, or of associate studies. A good example was the annual publication by Hulton Pres Ltd. between 1947 and 1957 of their *Readership Surveys* which provide a good deal of useful social and economic data. These have no ceased. A wealth of similar information is provided by similar studie by Odhams Press Ltd., and more recently the Institute of Practitioner in Advertising has published some general tables of interest, again a a by-product of readership surveys.[3] It is to be hoped that such publi cations will continue; to many they have proved most valuable.

Professor Devons's book, by and large, is confined to official statistics It is primarily designed for the student; provides him 'with a genera survey of the main British economic statistics—tells him where the can be found, what they mean, what problems arise in using and inter preting them and guides him in further reading and in consulting th original sources themselves.' But he hopes it will be useful to others including those in business who have to use published statistica information.

Space alone would clearly preclude Professor Devons from dealin in his modest-sized volume with all sectors of economic statistics. An the very process of continuing expansion and improvement in th publication of official statistics would mean that this type of book need continuous revision if it is to be kept up to date. In this respect, Si Harry Campion's paper serves a very useful purpose. But in the ligh of our earlier comments on what is required from data, Devons's boo serves admirably to remind the reader, in a number of ways, ho published statistics do not measure up at all precisely with what at fir blush they seem to measure. It is an understanding of the limitation of such data which are not obvious that helps to distinguish the pro fessional statistician from the layman.

In the main Professor Devons devotes attention to the well-establishe economic sources. These, too, are the sources most used in busines

He commences with population data. To many business statisticians the census of population will be in very frequent use, providing a comprehensive statistical picture of the nature of the population including, for example, considerable detail of local and regional populations. In between censuses some of this information is brought up to date by Quarterly and Annual Reports of the Registrar-General.

Generally speaking, we are well served by our population statistics. A user's guide to the censuses would probably need to devote far more attention than Professor Devons has been able to do to a number of aspects like household and housing statistics, especially as the role these play in economic development increases with a growing economy. In general, however, the more detailed one's use of census reports, the greater the need to consider the accuracy of census methods of population measurement and to examine definitions carefully—considerable care is, of course, always required anyway. And on the subject of definitions, Devons carefully draws attention throughout his book to what the figures cover, and often deals explicitly with what they do not cover. For example, there are three bases of estimating the size of the population in current use by the Registrar-General—the home population, civilian population, and total population. It is too easy to quote, without qualification, a figure for the population of England and Wales at, say, 44·9 million. This figure is the home population. But the size of the civilian population, which excludes armed forces of all nations in England and Wales, is 44·4 million, while a third figure in use is the total population, which includes British armed forces at home and abroad, but excludes any Commonwealth and Allied armed forces in the country who are included in the figure of the home population. The total population at the same date (mid-1957) was 45·0 million.

The need to make sure that one is using the same groups of people in tracing trends over time in population, national or local, is stressed in Professor Devons's book. Another not unimportant source of error in tracing changes in local populations is the effect of boundary changes. To those of us who have need to watch changes in the populations of local areas, one pitfall to be avoided is caused by the overflow of a local authority's housing programme into an adjoining local authority's area. What can happen is that significant increases are shown for a time in a Rural District's population owing to houses being built there by an adjoining Borough Council. Sooner or later, however, that part of the Rural District is officially absorbed into the area responsible for the housing development—and the population of the Rural District falls. Unfortunately, the cause of the changes can only emerge with the publication of boundary changes in the half-yearly reports of the Registrar-General, some time after the process has started.

While we are on the subject of population changes, another aspect of population statistics covered by Professor Devons should be mentioned. Usefully he discusses the important subject of population forecasting and the relative reliability of assumptions with regard to mortality and

fertility rates. The publication in recent years of long-term forecasts (projections) of the population of England and Wales by sex and quinquennial age groups by the Registrar-General has been welcomed by many; it has saved some of us a lot of time necessarily spent in earlier days in making our own population projections. What one would like now are similar projections for Scotland and the publication of regional projections for Great Britain. It is recognised, however, that this raises the difficult question of internal migration, on which little statistical information seems to be available.

Next in Professor Devons's review comes manpower data. A good deal of information on the occupational and industrial distribution of manpower is available in the Censuses of Population, Production, and Distribution. The main source for most purposes is the monthly *Ministry of Labour Gazette*, and it is to this that the analyst of business conditions most frequently turns. In this context, we are most often concerned with two categories of information, long-term trends and short-term changes.

With regard to long-term trends, Devons's notes on the comparability of pre-war, pre- and post-1948 employment series, though brief, are useful. But much more needs to be done in linking up and tracing changes in employment before and since the 1939–1945 war. This is one direction in which the Ministry of Labour and National Insurance could obviously help most instructively. Even more important is a clearer understanding of statistical factors influencing short-term changes in manpower data. Thus the seemingly straightforward count of unemployed, based on the lodgement of insurance cards by unemployed persons, may inadequately reflect the extent to which married women employed—now nearly 4 million in number—fail to register on losing a job. Another series of data which figures prominently in many economic commentaries is that of unfilled vacancies. But how completely this reflects changes in the labour situation remains to be assessed, as does the nature of the relationships, if any, between changes in vacancies and unemployment and other parameters of the labour situation. Much might be done, too, to elaborate the information published on earnings. Thus, in addition to publishing the average earnings twice yearly in different industries, could we not have complete distributions, as was the case from time to time, though very rarely, pre-war? Perhaps, too, we might some time have data by size of organisation. But less ambitiously, a start might well be made by an official attempt to adjust seasonally the vacancies data. Since this paper was first drafted, seasonally adjusted data on unfilled vacancies and unemployment in the form of proportions or total numbers employed have been given in the First Report of the Council on Prices, Productivity and Incomes (the Cohen Committee).[4] The adjustments have been made by the staff of the National Institute of Economic and Social Research.

Official statisticians are, by force of circumstance, cautious statisticians

and rightly so. It is this which probably explains their reluctance to undertake such hazardous exercises as seasonal adjustment to data spanning a limited number of years. What usually happens is that the unofficial statistician steps in and provides a service. This has occurred with the 'Vacancies Outstanding' data and, more importantly, with the official Index of Production. Today official seasonal adjustments of the Production Index for all industries covered and for manufacturing industries as a whole are published; perhaps we shall soon have similarly adjusted figures for the separate industries included in that Index.

So much attention is paid to the monthly Index of Industrial Production that the business economist is well advised to understand clearly the mechanics of production index construction. In this respect Professor Devons's discussion of the general problems of construction of such indices and his description of the current British index serves as a useful refresher, to be applied from time to time by the user. But what is still needed is a thorough review of the performance of the Index over recent years in the light of the Census of Production results for 1951 and 1954; perhaps this may come if the Index is revised to take account of the results of the 1954 census. In addition, many would doubtless find useful not only a sub-analysis of the large 'engineering, shipbuilding and electrical goods' group, to the need for which Devons alludes, but also the monthly publication of indices for such groups as 'durable goods' and 'non-durable goods' industries, as provided by the US Department of Commerce; annual Indices are already published in the *Annual Abstracts of Statistics*.

While on the subject of production indices, we owe a debt to the late A. A. Adams, who was responsible for the seasonally adjusted London and Cambridge Economic Service Index of Industrial Production, which filled a gap for some time. His paper on 'The Future of Production Statistics',[5] given to the Royal Statistical Society's First Conference on Economic Statistics in Cambridge in July 1956, just prior to his untimely death, deserves to be well known. In particular, his paper serves to remind the user of this Index of the effect of quality changes and the effect of services on the measure of industrial output. In an economy which buys an increasing amount of service of one type or another, be it packaging or convenience in application of the product concerned, an index relying mainly on changes in quantity of output could well be less than accurate.

A sector of economic activity about which we know much more today than we knew only five years ago is distribution. Thanks to the first official Census of Distribution, we now have a firm estimate *inter alia* of the number of shops in different categories of retail businesses and of different scales of business, the volume of wholesaling undertaken, and their regional dispersal. What we do not know, but should like to know from this Census, are the number of outlets for different types of commodities. Following the report of the Verdon Smith Committee on Censuses of Production and Distribution,[6] a sample census is being

undertaken this year in respect of 1957. It may be recalled that this Committee recommended that the Board of Trade should consider with its Advisory Committee the advisability of collecting information on outlets. This is not being done in the sample enquiry but it is to be hoped that the recommendation is not lost sight of. It is disappointing that in Devons's book so little space is devoted to the 1950 census; a more comprehensive review of the scope of the four reports published would be welcome.

Of greater value, particularly in short-term analysis of trading conditions, is the brief description by Professor Devons of the coverage of the Board of Trade's retail trade indices. But since the book was written, there have been important changes in the construction of these indices. The revised statistics now give a full coverage of retail trade; the panel of independent retailers, previously inadequate, has now been increased, and the index numbers shown are constructed on a basis of weighting according to sales as compared with the simple 'weighting' by number of returns previously used. Another major change has been the compilation of the indices on a 'kind of business' basis rather than a 'commodity basis' previously used.[7]

Like that of so many other official indices of this type, their value to the user depends largely on the amount of detailed sub-classification published. Although the revised indices are of considerable value in tracing month-by-month sales as a whole and by different types of retailer within various broad groups of merchandise, smallness of samples precludes assessment of, for example, the growth of multiple retailing in expanding fields, such as radio and electrical goods, in a way which is possible from these statistics in the case of food and clothing. For many engaged in business research, this growth of multiple retailing in a number of fields and the ability to measure and examine it is important. In due course, the sample surveys and Censuses of Distribution will enable the trends in recent years to be examined.

Statistical Spots and Symptoms

In this brief review of population, manpower, and production data we have been concerned mainly with the adequacy of published official sources for the purpose of formulating *statistical* experience on which to base judgements as to what has been, and is, happening, and why. When we turn to the later chapters of Professor Devons's book, on foreign trade and prices statistics, the possibility that figures available do not measure what they are believed to measure—or at least not very precisely—takes on additional significance. These series are news, the nature of which is of more industrial, financial, and economic significance than many of the other data we have just been considering.

Thus, considerable attention is given by economic commentators and economic analysts to the month-to-month position with respect to exports and imports and the apparent visible trade gap. But it is not

always appreciated that the value of trade recorded in the official statistics in any particular month relates to the export and import documents dealt with in that particular month by the Statistical Office of the Customs and Excise. There is scope for significant variations in the period elapsing before the export of goods and their inclusion in the month's returns owing to changes in the rate at which returns are made out and sent to the Statistical Office; and these changes have nothing to do with fluctuations in the actual volume of trade taking place. For this reason, Professor Devons points out, 'it is unwise to attach great importance to short-period changes in the trade figures from one month to the next' (page 142). Similar difficulties obtain in respect of the meaning of indices of price movements and of a concomitant index in wide circulation, that of the terms of trade.

Equally careful, if not even more careful, handling is required of Index Numbers of Retail Prices. The cost of whose 'basket of goods' does it measure? How well does it measure the change in its cost? Are the prices charged by 'cut-price stores', and the effect of 'coupon offers' with which the housewife is bombarded, taken into account—and should they be? These are some of a variety of questions which may be asked. Much is made available as to how the index is compiled, and the basic problems confronting the compiler of retail price indices are well known and do not require repetition here. But there is clearly room in a book dealing with British economic statistics for a fairly detailed discussion of the implications of the alternative solutions that can be adopted, tackling such problems as the complex of goods to be covered by the index, the weighting applied to the various components, and the base year to be adopted. In this sense Professor Devons has possibly found himself too much restricted by the limited size of his volume to do as much justice to this very important aspect of economic statistics as he might have wished. In any case it may be that the cynic would conclude, after a detailed study of these aspects of retail price index construction and consideration of their political impact, that suspension of publication of the index for, say, a year or two might considerably improve the economic climate.

The final chapter in Devons's book describes in its course the associated Indices of Wage Rates and Wage Earnings. Two of the aspects dealt with are worthy of particular note. One is the limited value of average earnings data for particular industries; data on variations within industry are clearly needed. The other is a useful brief description of the published pre-war series of wage rates. The reasons for the divergence between the Index of Wage Rates and that of Average Earnings is of particular significance in discussions of the causes and effects of inflation in post-war years. A number of factors have contributed to this divergence and it would be useful if their contribution could be assessed and the results published from time to time.

From wages, Professor Devons turns to other sources of income and their distribution, and thence to national income and social accounting.

This is a useful summary, but the professional economist will need to have by him when using the *National Income Blue Book* (and it is probably the most used of his sources) the official Guide[8] which became available after Professor Devons's book was written, although this, by now, requires revision to bring it up to date, and will presumably continue to require such revisions. Hence the need for the *Blue Book* each year to include notes which help the reader to keep up to date the description of sources, methods and definitions given in the Guide. One final point may be made before leaving the last chapter: there is a reference to the detailed input/output matrix for the year 1948, which the Board of Trade, in collaboration with the Department of Applied Economics, Cambridge, are compiling; it is to be hoped that its publication will not be much longer delayed.

The Filling of Gaps

There, then, are just some of the limitations of what might by and large be called well-established data. Now we must turn to more recent developments, most of which were not available when Professor Devons's book was prepared.

In *Economic Trends* of May 1957 the Government set out what they were doing to improve their 'out-of-date Bradshaw', to the deficiencies of which Mr Macmillan, while Chancellor of the Exchequer, had drawn attention. By May 1957 the quarterly surveys on fixed investment were well known, and it was announced that these were being supplemented by annual sample enquiries into the distributive and service trades. It was also announced that new quarterly figures on contracts accepted by the building and civil engineering industries were being gathered and that data on quarterly profits had begun to be collected from a sample of companies. These have, at the time of writing, not been published but the promised quarterly estimates of personal income have since appeared.

Other 'empty boxes' are being filled. The work in progress on data on stocks, on estimates of consumer expenditure by means of continuing quarterly family expenditure enquiries, and on industrial production measurement, are described in Sir Harry Campion's Presidential address to the Royal Statistical Society in November 1957. Also discussed in that paper is the scope of published financial statistics. One use would be a flow-of-funds matrix similar to the flow-of-goods-and-services matrix given by an input/output table. It must be some time before such integrated descriptions are available, but clearly the course of development in economic statistics is in this direction. Meanwhile there is still need for more work on the various areas of financial statistics like the analyses recently published by H. B. Rose and others on company accounts.[9, 10] Presumably a good deal of such work will become available with the publication, in due course, of the report of the Radcliffe Committee.

The Central Statistical Office publication *Economic Trends* deserves

a special mention in any review of 'gap-filling'. It is now quite customary for issues to contain brief, but usually very informative, articles on developments in various statistical series by Government Statisticians. The issue for January 1958, for example, describes the new quarterly estimates of national income and expenditure and, in particular, deals with the new quarterly series on personal income. These texts, by the way, often include data not usually covered by the published tables. With *Economic Trends* we seem to be moving in the direction of the American *Survey of Current Business*; that, however, usefully embodies the results of research by Government Statisticians and Economists and is more analytical. Perhaps we can look forward to further steps in this direction with *Economic Trends*.

Filling the statistical gaps must be a job of increasing difficulty. And we must beware, as Sir Harry Campion has put it, of inserting figures into tables 'which are really only papering over the vacuum of knowledge'. The issue ultimately is one of comparative advantage; the structure of statistical data is the framework of experience within which the business economist or statistician performs his professional duties. He must play his part in improving and developing data; the expert use he makes of it, and the extent to which he understands it, determine the value of the advice he gives and ultimately his professional status.

REFERENCES

[1] Devons, E. (1956). *An Introduction to British Economic Statistics*. Cambridge University Press. Pp. vii + 255. 22s. 6d.

[2] Heyworth, Sir Geoffrey. 'The use of statistics in business', p. 3, *J. R. Statist. Soc.*, A, **113**, 1.

[3] Institute of Practitioners in Advertising (1957). *The National Readership Surveys*. An explanatory manual prepared by Mark Abrams. I.P.A., London.

[4] *First Report of the Council on Prices, Productivity and Incomes* (1958). HMSO, London.

[5] Adams, A. A. (1956). 'The future of production statistics', unpublished; available in Royal Statistical Society Library.

[6] *Report of the Committee on the Censuses of Production and Distribution* (1954). Cmd. 9276. HMSO, London.

[7] 'Improved monthly statistics of retail sales', *Board of Trade Journal*, 5th May, 1956. HMSO, London.

[8] *National Income Statistics. Sources and Methods* (1956). Studies in Official Statistics No. 3. Central Statistical Office. HMSO, London.

[9] Rose, H. B. (1957). 'Monetary policy and the capital market, 1955–1956', *Econ. J.*, **67**, 397.

[10] Rose, H. B. (1958). 'The finance of inflation', *Lloyds Bank Review*, Jan. 1958, 1.

SIMPLE MORTALITY RATES

SIMON BROADBENT

The United Glass Bottle Manufacturers Ltd.

Dr Broadbent here shows how simple forms of expression fo the mortality rate may be used in the study of the failure c articles in service. He takes as illustrative examples th behaviour in service of milk bottles and tumblers, and demon strates the use of techniques of varying degrees of sophistication

The Distributions

We here consider a process in which manufactured units are put int service and eventually fail. The units might be, for example, electroni valves or nylon stockings or milk bottles; in these cases failures mea the burning out of a valve, the laddering of a stocking or the breakin of a bottle. The time in service before a unit fails cannot be predicte exactly, but is subject to random variations.

This situation was first recognised in man himself, and it is therefor usual to talk of the age of a unit at death, and of mortality distribution and rates. Specifically we shall measure the age of a unit, t, from th time it was first put into service, and write $f(t)$ for the *mortality distribu tion.* Then $f(t)dt$ is the probability that a unit will die in the interva $(t, t + dt)$. The probability that a unit will still be alive at age t i written $s(t)$, the *survival function*, and clearly

$$s(t) = 1 - \int_0^t f(t)dt$$

If we know a unit is alive at age t, the probability that it will die befor $t + dt$ is $m(t)dt$, where $m(t)$ is the *mortality rate* at age t, and

$$m(t) = f(t)/s(t)$$

The last equation is a functional relation, from which we can easil deduce that, if

$$M(t) = \int m(t)dt + C$$

where C is chosen so that $M(0) = 0$, then

$$\left.\begin{aligned} s(t) &= \exp[-M(t)] \\ \text{and} \qquad f(t) &= m(t)\exp[-M(t)] \end{aligned}\right\} \quad \ldots . (1$$

It is the object of this paper to discuss several mortality functions in which the mortality rate $m(t)$, rather than the mortality function, takes a simple explicit form and has an obvious intuitive meaning. These functions are suitable for the description of a very wide range of processes met in practice. We shall suppose that there is a reasonably large amount of data available, and so we shall not discuss small-sample estimation or the planning of life tests.

We shall consider mortality rates which are either constant or vary as a linear function of age, as the age raised to a power, or as an exponential function of age. These processes are set out in Table I, from which the probabilities of living to each age and the mortality distributions may be written down immediately using equations (1).

TABLE I

Mortality processes

Name of Process	Mortality Rate $m(t)$	$M(t)$
Constant-exponential ..	a	at
Linear-exponential ..	$a + 2bt$	$at + bt^2$
Stange[10]	$a^b bt^{b-1}$	$(at)^b$
Gompertz[5]	ae^{bt}	$\frac{a}{b}(e^{bt} - 1)$

It will be noted that the constant-exponential distribution is a particular case of each of the others. The most familiar and exact example of this distribution is found in the process of radioactive decay, but many other processes are well represented by this model.

There is a generalisation of the linear-exponential and Stange distributions, obtained when the mortality rate is zero up to age T, and is given by $m(t - T)$ for $t > T$. For continuity in the linear-exponential case we should require $a = 0$. For the Stange case the resultant distribution has been discussed by Weibull[11] and Kao[6] and is known as the Weibull distribution. There is a generalisation of the Gompertz distribution in which $m(t) = ae^{bt} + c$, known as the Makeham[7] distribution. The last two distributions have been much used in smoothing census data to produce mortality tables.

There will usually be two objects in investigating a mortality process aid in fitting a mortality distribution. The first is to decide what form the mortality rate takes, and if possible to relate this to the technology involved. The second is to estimate the parameters of the distribution, and in particular to calculate the mean life. The means of each of the distributions in Table I are listed in Table II. If the mortality rate is zero until age T, as in Weibull's generalisation of the Stange distribution, the mean is of course increased by T.

TABLE II

Means of mortality distributions

Name of Process	Expected Life, or Mean Age at Death
Constant-exponential ..	$1/a$
Linear-exponential ..	$-a/2b + b^{-\frac{1}{2}}\mathrm{L}(a^2/4b)$ where $\mathrm{L}(x) = \mathrm{e}^x \int_x^\infty y^{\frac{1}{2}}\mathrm{e}^{-y}\mathrm{d}y$ and is tabulated in Table III
Stange	$\frac{1}{a}\Gamma\left(\frac{1+b}{b}\right)$ where $\Gamma(x)$ is the gamma function
Gompertz	$\mathrm{G}\left(\frac{a}{b}\right)/a$ where $\mathrm{G}(x) = \mathrm{e}^x \int_x^\infty y^{-1}\mathrm{e}^{-y}\mathrm{d}y$ and is tabulated in Table III

NOTE. The integral in $\mathrm{L}(x)$ can be derived from the incomplete gamma integral[9]; th integral in $\mathrm{G}(x)$ is the logarithmic integral, tabulated, for example, by Jahnke and Emde

TABLE III

$$\mathrm{L}(x) = \mathrm{e}^x \int_x^\infty y^{\frac{1}{2}}\mathrm{e}^{-y}\mathrm{d}y$$

$$\mathrm{G}(x) = \mathrm{e}^x \int_x^\infty y^{-1}\mathrm{e}^{-y}\mathrm{d}y$$

x	$\mathrm{L}(x)$	$\mathrm{G}(x)$
0	0·886	∞
0·1	0·951	2·015
0·2	1·012	1·493
0·3	1·067	1·223
0·4	1·119	1·048
0·5	1·168	0·923
0·6	1·214	0·828
0·7	1·258	0·753
0·8	1·300	0·691
0·9	1·341	0·640
1	1·381	0·596
2	1·712	0·361
3	1·987	0·262

We now turn to the information available when we study mortality process. We shall assume that the units being studied for a homogeneous population with an unknown but common mortali distribution.

Example 1: Assuming Constant Mortality Rate

In the simplest case we may not have enough information to determine anything about the mortality distribution except some average value and we therefore cannot expect to discover which of the possible processes is actually occurring. For example, we may be studying the lives of milk bottles that are filled in a dairy, circulated to customers, and returned empty to the dairy. Suppose we have available the first two columns only of Table IV. We do not at this stage assume anything to be known about the age composition of the present population of bottles; we know only its total size (in this case the dairy fills 90 000 bottles a day and a trip lasts on the average 5 days) and the numbers of new bottles purchased to replace those lost or broken.

TABLE IV

Milk bottles put into service in each period of three months. Unit: 10 000 *bottles*

r	b_r	n_r
0	6	—
1	20	16·3
2	12	8·6
3	17	6·6
4	15	5·4
5	15	2·7
6	10	1·7
7	18	1·3
8	13	0·7
9	23	0·8
10	11	0·1
11	6	—
12	17	—
13	11	—

NOTE

b_r is the number put into service between $(3r - 1\frac{1}{2})$ and $(3r + 1\frac{1}{2})$ months previously (b_0 is the number less than $1\frac{1}{2}$ months previously).

n_r is the number of the b_r now surviving, estimated from a sample.

The first calculation we may perform is to find a simple average: in three months 143 000 new bottles are put into service (the average of the b_r) and 8 190 000 trips are made (90 000 bottles a day for 91 days). The quotient, 57, is the average number of trips per bottle. The reciprocal of 57, or 0·0175, is the probability that a bottle will be broken in any one trip. This calculation neglects any change that may be taking place in the size of the population; it assumes that steady conditions have been reached. This assumption is not needed for the following procedure, which involves more calculation.

Let p be the probability that a bottle survives three months. Since a bottle makes on average 91/5 trips in this time, and we have estimated 0·9825 as the probability of surviving each trip, we derive as a first estimate of p: $(0{\cdot}9825)^{91/5}$ or 0·7242.

Now the expected number of bottles surviving equals approximatel

$$\sum_{r=0}^{\infty} b_r p^r$$

where b_r bottles have been subjected to approximately r periods of risk in each of which the probability of survival is p.

Equating this expression to the actual number circulating, 450 00 (i.e. 5 times 90 000), we obtain an equation for p which can be solve by Newton's method of successive approximation or by graphica interpolation. The differentiation of

$$\sum_{r=0}^{\infty} b_r p^r$$

gives another expression easily calculated, since tables of p^r and of rp^{r-} can be quickly constructed for the relevant p. The sum over large value of r can either be neglected (p^r being small) or approximated, sinc

$$\sum_{n}^{\infty} b_r p^r$$

is $bp^n/(1 - p)$, where b is an average value of the b_r.

In this example the next approximation to p, by Newton's method is 0·7130 and the third and subsequent approximations 0·7126. Th average number of trips and probability of failure in any trip are now estimated as 54 and 0·0185 respectively. These are close to the origina estimates because the rate of supply (the b_r of Table IV) does not show any trend over the period considered. However, where the size o population is not steady, this method can give results quite differen from, and of course more accurate than, those of the previous crud method.

Example 2: Given Survival Data

We continue the previous example, of bottles circulating in a dairy without the assumption of a constant mortality rate and with addi tional data. The last column of Table IV is now assumed to be avai able. This supposes that a sample of the surviving bottles has bee taken, and that from it the numbers of each age group now survivin have been estimated. From the survival data and from the numbe of each age group put into service we obtain $s_r = n_r/b_r$, the proportio of bottles which has survived $3r$ months. These are direct estimates o the values of $s(t)$, the survival function. The precision of the estimat depends on the size of sample taken and the effect of grouping, and wi not be discussed here.

It follows from equations (1) that $-\log s(t)$ is $M(t)$. From Table it follows that $M(t)/t$ is constant if the mortality process is constan exponential and linear if the process is linear-exponential. We therefo

lot $-(\log s_r)/r$ against r, as shown in Fig. 1. This graph may be used o decide whether either of these processes is relevant, and to estimate he parameters.

In the dairy example, a linear-exponential process fits the data well, n which

$$s(t) = \exp[-(0{\cdot}16t + 0{\cdot}03t^2)]$$

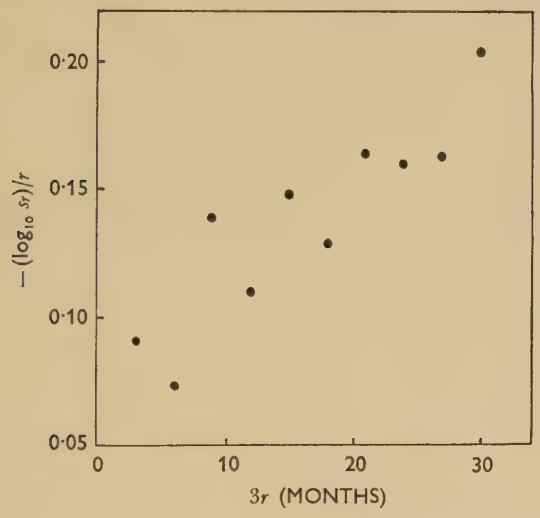


FIG. 1. Dairy example: s_r is the proportion of bottles surviving $3r$ months.

hen t is measured in units of 3 months. The mortality rate is

$$0{\cdot}16 + 0{\cdot}06t$$

The probability that a bottle has survived to time t is $s(t)$; the robability that it survives to time $(t + 5/91)$, which is one trip later, s there are 91/5 trips per unit of time, is $s(t + 5/91)$. The latter is also $(t)P(t)$, where $P(t)$ is the probability that the bottle survives a trip tarted at age t. Thus

$$P(t) = s\left(t + \frac{5}{91}\right)/s(t)$$

nd the probability of failure on this trip is

$$\left\{s(t) - s\left(t + \frac{5}{91}\right)\right\}/s(t)$$

ow $s(t) = \exp[-M(t)]$, and a good approximation will be

$$s\left(t + \frac{5}{91}\right) \simeq s(t) - \frac{5}{91}m(t)s(t)$$

It follows that the probability of failure in any trip is approximately $5m(t)/91$, or

$$0{\cdot}009 + 0{\cdot}003t$$

This compares with the previous (constant) estimate of 0·0185. The two are equal at about $t = 3$, or 9 months. There is evidently some cause, such as scuffing and surface weakening from repeated thermal shocks in washing, which causes older bottles to be more likely to be removed from service in any one trip than new bottles. The mean life of bottles is now estimated, using Tables II and III, as 59 trips. This result is close to the previous estimates but is, of course, based on quite different assumptions. It is only because this process has reached a steady state that a similar estimate has been obtained.

The plot of $-(\log s_r)/r$ against r will not be constant or linear in all processes: if the mortality process follows Stange's model the relation will be concave decreasing for $b < 1$, convex increasing for $1 < b < 2$ and concave increasing for $b > 2$. (In this and similar contexts, 'convex' means that the graph would appear convex if viewed from above with a corresponding meaning for concave.) If one of these cases is indicated $\log(-\log s_r)$ should be plotted against $\log r$. If this is linear, estimates of a and b in the Stange process can be obtained. If on the other hand $\log(-\log s_r)$ is large and negative for small r, and then convex against $\log r$, Weibull's generalisation is indicated, T may be estimated from the age the first death occurs, and the Stange process from time T studied.

The Gompertz or Makeham processes are indicated when $-(\log s_r)/$ is found to increase exponentially with r. The Gompertz process is indicated if the logarithms of the first differences of $\log s_r$ (taken at equal intervals) are linear against r. The Makeham process is indicated if the preceding plot is convex, but the logarithms of the second differences of $\log s_r$ are linear.

More sophisticated methods of estimating parameters and for testing which process obtains (usually by a χ^2 test) will not be discussed here. For example, the Makeham mortality distribution was one of the first curves to be fitted by moments (Pearson[8]); Freudenthal and Gumbel have recently fitted the Weibull distribution by moments; Kao[6] discussed fitting the Stange distribution by least squares and by maximum likelihood.

Example 3: Given Mortality Rate Data

The example given to illustrate this situation is quoted from Brown and Flood[1]. Toughened tumblers circulating in a cafeteria were studied by noting for successive periods, each of five or ten weeks, the number exposed to risk, N_r, and the number broken, n_r. The population of tumblers under investigation was not made up to replace breakages, so that N_r continually decreases. However, other tumblers were added to keep the total number circulating, and the use made

ach tumbler in the population, constant. The proportion broken in ne rth period, $m_r = n_r/\mathcal{N}_r$, estimates the mortality rate during this eriod. The data are reproduced in Table V.

Brown and Flood give interesting technological reasons for supposing ne appropriate mortality distribution to be similar to a gamma distribu- on, and fit a gamma process to the data by a rather laborious inimum-χ^2 procedure. They obtain a very good fit ($\chi^2_{11} = 7{\cdot}0$), hich is shown by the calculated values n_r' in Table V. The corre- ponding expression for the mortality rate is complex. Glathart and reston[4] have derived a similar model for the breaking of glassware, hich they have fitted satisfactorily by moments. Their expression for ne mortality rate is approximated by $\mathrm{m}(t) = a(1 - \mathrm{e}^{-bt})$.

TABLE V

Tumbler mortality: original data collected by Brown and Flood[1]

Period r	$\mathcal{N}_r$	n_r	n_r'	n_r''
1	241	7	6·75	10·48
2	223	11	10·93	11·55
3	208	10	12·48	10·86
4	188	9	12·78	12·69
5	170	17	12·58	12·46
6	146	10	11·39	11·48
7	132	13	10·82	11·01
8	110	9	9·35	9·65
9	94	6	8·18	8·62
10	86	9	7·74	8·21
11	73	5	6·79	7·22
12 & 13	60	14	10·86	11·70
14 & 15	45	10	8·28	9·35

NOTE

$_r$ is the number of tumblers submitted to risk in the rth period of 5 weeks, of which n_r were broken.

' is the fit by a gamma process, $\chi^2_{11} = 7{\cdot}0$.

'' is the fit by a Stange process, $\chi^2_{11} = 5{\cdot}4$.

We now approach Brown and Flood's data rather differently, hoping find a simple expression for an appropriate mortality rate $\mathrm{m}(t)$, ather than thinking in terms of a mortality distribution which is ormal, lognormal, or gamma, and to obtain in this way a more direct sight into the process. We therefore attempt to fit one of the pro- sses of Table I.

The first step is to plot m_r against r, as this will be constant or linear a constant-exponential or linear-exponential process. The para- eters a and $2b$ can be estimated from the graph. But in a Stange ocess m_r will be concave decreasing if $b < 1$, convex increasing if $< b < 2$, and concave increasing if $b > 2$. In a Gompertz process will increase exponentially, which is of course difficult to distinguish om a concave Stange process.

If a Stange process is indicated, plot log m_r against log r. This will

be linear, and estimates of $a^b b$ and $(b - 1)$ can be obtained. If it convex a Weibull process is indicated, and T may be estimated from th age at which the first death occurs. If, however, log m_r is concav against log r a Gompertz or Makeham process is indicated, and log m should be plotted against r. If this is linear, estimates of the parameter of the Gompertz process may be obtained, but if concave, a Makehar process is indicated.

In the tumbler example, Fig. 2 shows the relation between log m_r an log r. The Stange process evidently gives a good fit to the data, an

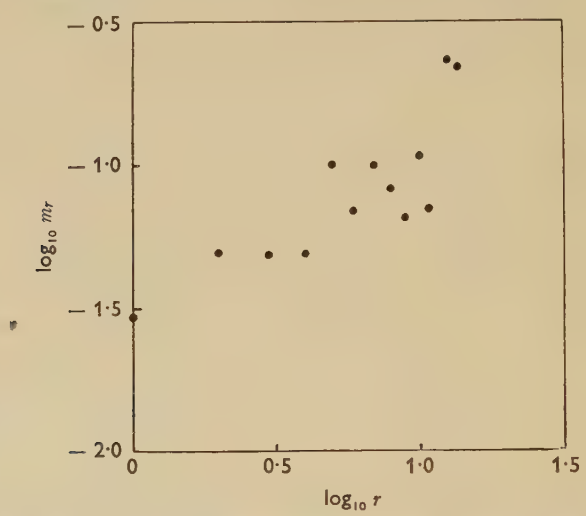


FIG. 2. Tumbler example: m_r is the proportion of tumblers broken in the rth period of fi weeks.

this is borne out by the calculated entries n_r'' in Table V. These we obtained with $a = 0{\cdot}077$ and $b = 1{\cdot}38$, and give a χ^2_{11} of 5·4, rath smaller than the χ^2 obtained by Brown and Flood. The mean lif obtained from Tables II and III, is calculated to be 59 weeks, in clo agreement with Brown and Flood's estimate. In fact the two processe Stange's and gamma, have mortality rates and distributions identic for all practical purposes in this example. Stange's process is, howeve more easily fitted.

We conclude that in this example the mortality rate is $0{\cdot}040r^{0{\cdot}3}$ i.e. this is the probability of breakage during the rth period of 5 week given that a tumbler has survived to the beginning of the period.

Finally, suppose that data from the mortality distribution itself a available, i.e. we have the actual ages at death of a sample, possib truncated. It is practicable to fit certain distributions to this dat However, for the processes we are considering, it is easier to convert t data to mortality rates by grouping the data. That is, if the samp consists of n ages at death and nm_r of these ages fall into the rth perio we may return to the situation considered at the beginning of t example.

REFERENCES

Brown, G. W. and Flood, M. M. (1947). 'Tumbler mortality', *J. Amer. Statist. Ass.*, **42**, 562.

Freudenthal, A. M. and Gumbel, E. J. (1954). 'Minimum life in fatigue', *J. Amer. Statist. Ass.*, **49**, 575.

Jahnke, E. and Emde, F. (1953). *Tables of Functions.* Dover Publications, New York.

Glathart, J. L. and Preston, F. W. (1948). 'Theory of behavior of glassware in service', *J. Amer. Ceram. Soc.*, **31**, 153.

Gompertz, B. (1825). 'On the nature of the function expressive of the law of human mortality and on the new mode of determining the value of life contingencies', *Phil. Trans.*, 513.

Kao, J. H. K. (1956). 'The Weibull distribution in life-testing of electron tubes', *J. Amer. Statist. Ass.*, **51**, 514.

Makeham, W. M. (1889). 'On the further development of Gompertz's law', *J. Inst. Actu.*, **28**, 152.

Pearson, K. (1902). 'On the systematic fitting of curves to observations and measurements', *Biometrika*, **1**, 265.

Pearson, K. (1922). *Tables of the Incomplete Gamma Function.* Cambridge University Press.

Stange, K. (1955). 'Zur Ermittelung der Abgangslinie für wirtschaftliche und technische Gesamtheiten', *MittBl. für Math. Statist.*, **7**, 113.

Weibull, W. (1951). 'A statistical distribution function of wide applicability', *J. Appl. Mech.*, **18**, 293.

MARGINAL COST*

ANDREW MUIR
J. Bibby and Sons Ltd.

In this article Mr Muir defines the term 'marginal cost', explain a method of estimating these costs, and gives an example of it use. The treatment of the subject is not claimed to be complete but because of the apparent apathy that exists on margina costing it is intended to be somewhat provocative.

Introduction

Probably the most useful statistical technique that can be employe in the field of accountancy is the use of regression coefficients as esti mates of marginal costs; it has not, however, been generally accepte by the accountants themselves as a practical method. With the in creasing demand for accurate estimates of marginal costs as data fo problems of optimisation, traditional methods of estimating these cost are no longer adequate and it is the purpose of this short paper t redefine the term marginal cost, to demonstrate the method of estima tion, and to give an example of its use.

The Relation between Activity and Expense

Before actually defining the term marginal cost we shall examin some general notions on the relation between activity and expenditure For convenience we shall consider a single machine producing only on type of commodity; it will not be difficult to generalise this at a late stage to a number of production units making many different product or to a concern where the activity is not necessarily one of production

If we denote the production over a constant length of time by x an the expenditure by y, then from purely logical considerations we shoul expect the function $y = f(x)$ to increase with x, otherwise we should b able to make something for nothing. Furthermore, the value of y a $x = 0$ will be greater than zero by an amount which we shall call a this being the fixed expenditure during the period on housing an maintaining the machine. The three basic types of relations are show graphically in Fig. 1.

The cost of producing a single unit is known as the *unit cost* and i given by the total expenses divided by the total production, i.e. b $f(x)/x$. This cost is not independent of the rate or level of productio and is therefore an undesirable statistic to use for most purposes. It ha a further disadvantage in that it includes the *fixed expenses*, a, or wha

* Based on a paper given at a Conference of the Industrial Applications Section of th Royal Statistical Society held in Nottingham, July 1956.

ight be better termed the *inevitable expenses*. If these expenses are ncluded in costs quoted to management they can be most misleading, s the choice between a number of courses of action should be made n the extra expense involved and not on the expense to which one is ommitted whatever action is taken.

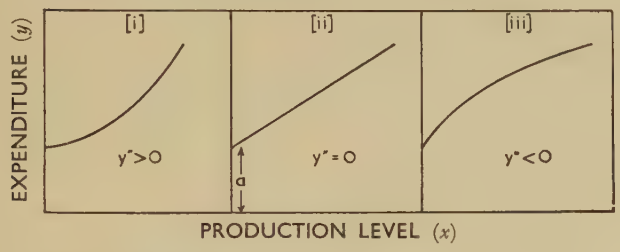


FIG. 1. Expenditure–production relations: three basic types.

If we consider now the cost given by the variable expenses (total xpense minus fixed expense) divided by the production it will be seen at in the linear case this is given by $(y - a)/x = b$, and is therefore ndependent of production levels. Now it fortunately happens that in ost cases it is this linear relationship that holds and it is this cost hich will be seen later to be the marginal cost.

he Conventional Cost Structures

It is customary among cost accountants to divide the various items expense of a production unit into two groups, which we shall call the rect charges and the indirect or overhead charges. The direct charges e defined[1] as those expenses which can be allocated to the particular oduction unit or machine we are considering. The indirect charges e those expenses which cannot be allocated but which are apportioned n some arbitrary basis to the different cost centres. Examples of some pical items of expense under the two groups are:

Direct Charges	*Overhead Charges*
Raw materials	Administration
Wages	Depreciation
Salaries	Rates
Services (electricity, gas, water, etc.)	Rents
Fuel oil	Insurance
Repairs and maintenance	Pensions

We should not expect the overhead charges to vary with output ovided that we restrict ourselves to the capacity of the plant; and nversely we should not expect direct charges for a particular process vary *directly* with output, because in almost every item there is a fixed ement of cost. A simple example of this would be the electricity bill, hich may include a fixed charge irrespective of the units consumed.

If we assume that a linear relationship holds between expenses an output then we can construct a simple diagram of the cost structure a different levels of production (Fig. 2).

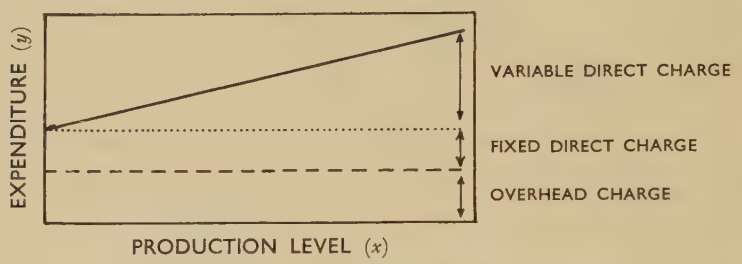


FIG. 2. A simple type of cost structure with linear relationship between expenditure an output.

If the equation of the line is $y = a + bx$, where a is the overhea charge plus the fixed element of the direct charge, then the slope of th line, b, is the marginal cost.

A Definition of Marginal Cost

It is unfortunate, in my opinion, that the Institute of the Cost an Works Accountants have adopted[1] as their definition of marginal cos

'Prime cost plus variable overhead',

where the prime cost is defined as 'the aggregate of direct materia cost, direct wages, and direct expense'. Apart from the fact that n indication is given as to how variable overheads are separated fro fixed overheads, the definition fails to single out the purely variabl elements of the direct expenses. It would appear that the definition one of convenience because of the impossibility of isolating the variabl portion of expenses by conventional methods of analysis. It woul therefore appear that the first essential is to redefine marginal cost. have already proposed[2] the following definition:

'The rate of increase of expense with respect to production.'

In more precise terms, if we denote the total expenditure ov periods of specified length by y and the output of different products b the variables $x_1, x_2, \ldots, x_n$, then the relationship will be of the for

$$y = f(x_1, x_2, \ldots, x_n) \qquad \ldots . (1$$

and the marginal costs of producing the different commodities will b given by the partial differentials

$$\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_n}$$

In the particular case where a linear relation holds, i.e.

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n \qquad \ldots . (2$$

ien the marginal costs are given by

$$b_1, b_2, \ldots, b_n$$

inearity

In every case I have come across (see for example references 3–5) ie relation between activity and expense has not departed significantly om the linear form. This fact is apparently offensive to the econo- ists, who consider the straight line as a very particular case and un- kely to hold in practice. Whether the relations are in fact linear, or st appear to be so over the range considered, is an academic one. he main point is that it is possible to test for departures from linearity, d if this does occur then marginal cost does not have the property of dependence of output but can still be estimated by regression analysis. or the remainder of this paper I shall be mainly concerned with the iear relation.

stimating Marginal Costs

The estimation of the marginal costs now follows quite simply from e definitions and is, of course, the normal procedure of least squares. we assume that errors occur only in the measurement of the expense, and not in the measurement of activity, as would seem to be the case, en least squares will provide the best estimates of our marginal costs. hese estimates will still be valid even if we have unwittingly ignored riables on which y is dependent. Thus in the simplest of cases:

$$y = a + bx \qquad \ldots.(4)$$

e best estimate of the marginal cost is given by

$$b = \Sigma(x - \bar{x})(y - \bar{y})/\Sigma(x - \bar{x})^2 \qquad \ldots.(5)$$

There are two methods of approach in calculating these marginal sts—the historical and the experimental methods. The historical ethod consists of extracting past data for the process under considera- on and using these data in the regression analysis. The precision of the sulting estimates will be increased by the number of periods available r analysis. As the marginal cost will generally be required to repre- nt present conditions it is preferable not to delve too far back into the dgers. Sometimes a compromise can be reached by renovating old ta, for example by adjusting data on wages to take into account anges in wage rates. The experimental method will consist in erating the production unit at different (determined) levels of pro- ction for each period and measuring the resulting expense.

Example on Steam Raising

At our Liverpool factory we have two sources of steam supply, our n boiler-house and a nearby power station. The question was raised

that in the event of our introducing new plant and either source c steam being available for use, would it be more economical to raise th extra steam in our own boiler-house or should we purchase it? As w were already committed to the fixed expenses on our boiler-house an payment of a fixed charge to the British Electricity Authority, it wa merely necessary to compare the marginal cost of our own steam wit the running charge for the purchased steam.

The data for the direct expenses of running our own boiler-house fo the years 1954 and 1955 are shown in Table I together with the outpu The periods are of four or five weeks but the five-weekly periods hav been corrected to four-weekly periods by multiplying by 0·8. It wa found to be more accurate to use the tonnage of coal used and multipl by the current price rather than to use the cost of the coal used.

TABLE I

Boiler-house running expense data

Year	Period No.	Wages (£)	Services (£)	Coal (tons)	Steam Produced (units)
1954	1	947	337	1 440	19 428
	2	1 059	350	1 480	20 831
	3	1 061	361	1 430	20 942
	4	978	290	1 200	16 544
	5	1 065	318	1 060	14 918
	6	897	283	990	13 776
	7	865	227	760	10 278
	8	822	142	450	4 572
	9	645	93	250	2 736
	10	595	135	450	5 521
	11	774	220	930	12 605
	12	776	186	820	11 922
1955	1	898	281	1 047	12 682
	2	847	221	714	8 542
	3	846	231	706	8 555
	4	815	213	570	6 966
	5	798	178	485	5 070
	6	774	96	150	1 364
	7	575	70	117	646
	8	629	87	124	304
	9	634	106	130	880
	10	690	112	209	1 347
	11	661	121	187	771
	12	625	89	215	956
Total		19 276	4 747	15 914	202 156
Mean		803	198	663	8 423

The expenditure on repairs and maintenance was examined, bu there was no statistically significant correlation with output.

The estimate of marginal cost has been broken down (see Table II into three components so as to demonstrate the relationship of thre different types of expenditure with production, and the correspondin

aphs are attached to this paper (Figs. 3–5). The regression coefficient r coal has been multiplied by £4 8s., the average price of coal at the ne, to give the marginal cost.

TABLE II

Marginal cost data

Item	Marginal Cost (£)	Variance of Marginal Cost	Standard Error of Marginal Cost
Wages	0·019 55	$4{\cdot}56 \times 10^{-6}$	$2{\cdot}1 \times 10^{-3}$
Services ..	0·013 13	$0{\cdot}56 \times 10^{-6}$	$0{\cdot}7 \times 10^{-3}$
Coal	0·289 39	$32{\cdot}91 \times 10^{-6}$	$5{\cdot}7 \times 10^{-3}$
Total	0·322	38×10^{-6}	$6{\cdot}2 \times 10^{-3}$

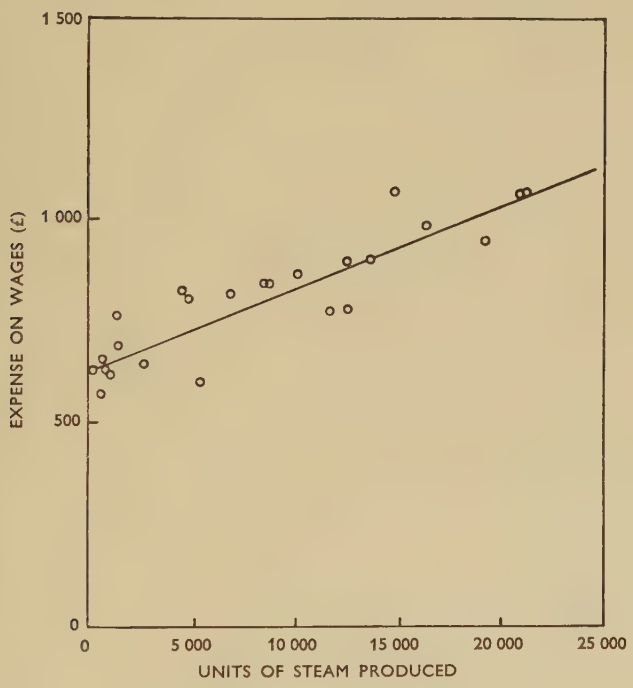


Fig. 3. Wages cost for production of steam.

The total marginal cost will be seen to be approximately 6s. 5d. ± . per unit of steam. Now the purchased steam was paid for by both ixed monthly charge and a marginal charge which took into account e price of coal. The marginal charge for 1955 averaged £0·342, or 10d. per unit. In order to test whether there is a significant dif- ence between the two charges we calculate

$$t = \frac{0{\cdot}342 - 0{\cdot}322}{0{\cdot}0062} = \frac{0{\cdot}020}{0{\cdot}0062} = 3{\cdot}23$$

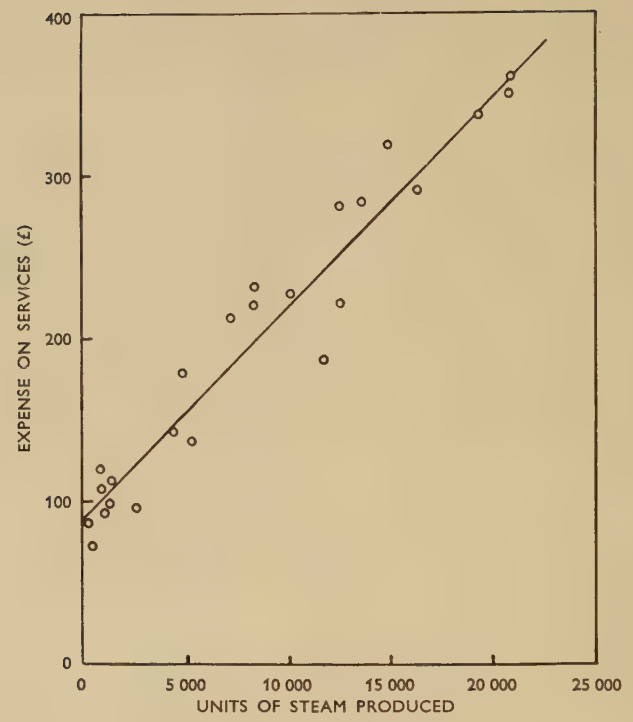


FIG. 4. Expenditure on services for production of steam.

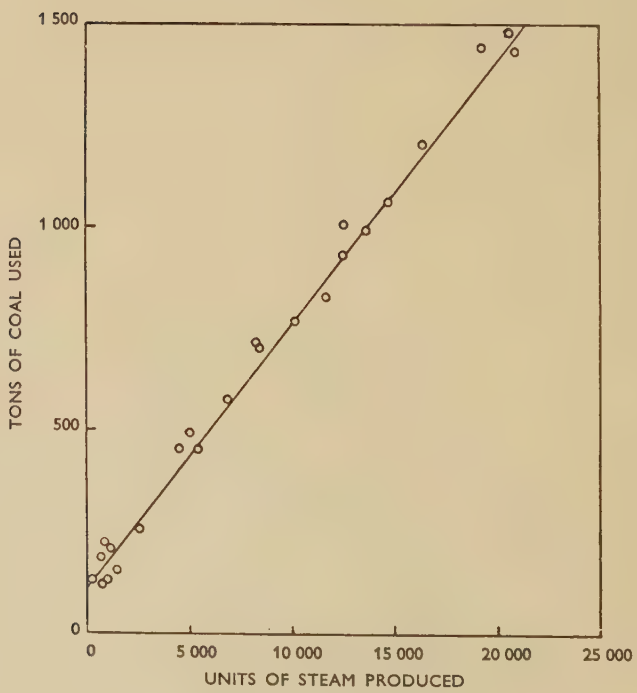


FIG. 5. Tons of coal used for production of steam.

is is significant at the 1% level for 22 d.f., and we conclude that our n steam can be produced (marginally) more cheaply than we can rchase it; hence if there is a choice of either source of steam we ould settle for our own.

If we had calculated the marginal cost according to the cost accoun-nts' definition, it would have been given by the average of the three ime costs (wages, services, and materials) divided by the average oduction, i.e. £3 918/8 423 = £0·465 or 9s. 4d. per unit of steam, d an incorrect decision might have been made.

The extension to several independent variables poses no further tistical problems and the estimation of the coefficients is similarly ade using the least-squares principle. It is interesting to note that it not possible to estimate multiple marginal costs by conventional counting methods.

nclusion

In this paper I have tried to demonstrate that there is a real need r a general adoption of the use of marginal costs and for their estima-n by statistical methods. This approach was first advocated by ilip Lyle,[6] and anyone who wishes to proceed further could do no tter than by referring to his book. There would appear to be too uch effort directed in the allocation of overheads to see whether each partment, or cost centre, is making a profit and not enough in obtain-g adequate information for management. Whether a single depart-nt, or production unit, makes a profit or loss seems to me to be ally irrelevant. What is of importance is whether each unit is being n in such a manner that the total profit of the concern as a whole is aximised. The statistical methods of estimating the marginal costs e relatively simple and could be made a matter of routine; if generally opted by the accountant they should enable him to provide the anager with the data he needs most.

REFERENCES

erminology of Cost Accountancy (1952). The Institute of Cost and Works Accountants, London.

he Cost Accountant, August 1956.

EAN, J. (1941). 'Statistical cost functions of a hosiery mill', *University of Chicago, Studies in Business Administration*, **11**, No. 4, 19.

NTEMA, T. O. (1940). 'Steel prices, volume and cost', *United States Steel Corporation T.N.E.C. Papers*, **1**, 231.

LIE, K. H. and EZEKIEL, M. (1941). 'Cost functions for the steel industry', *J. Amer. Statist. Ass.*, **36**, 777.

LE, P. (1946). *Regression Analysis of Production Costs and Factory Operations*. Third edition, revised by L. H. C. TIPPETT, 1957. Oliver and Boyd, Edinburgh and London.

A SIMPLE APPROACH TO THE MINIMISATION OF STORE-SIZE

NORMAN T. J. BAILEY
Unit of Biometry, 7 Keble Road, Oxford

In this article Mr Bailey describes an approach to th problems of checking the performance of existing stores an deciding the size of store needed in new buildings. Since n extensive mathematical knowledge is required in the use of th methods recommended, we hope that some readers may b encouraged to try them out and discover whether they ar adequate in practice.

Introduction

In recent years there has been a good deal of work on the mathe matical theory of storage systems, including such topics as inventor policy and the theory of dams (see, for example, the bibliography i Ackoff,[1] and the papers by Gani[2] and D. G. Kendall[3] in the 1957 Roya Statistical Society Symposium on Storage, where many references ar given). Useful practical applications can often be made in appropriat circumstances. Take, for instance, a commercial warehouse, where th frequency distribution of demand, the time-lag for ordering fresh stock and the losses incurred when customers' orders cannot be met or whe stocks have to be held in excess of warehouse capacity, are all know with sufficient accuracy. Then it is possible to compute an optima inventory policy that minimises losses. In such a situation the genera conditions are supposed known, and we try to find the most economi method of operation. Alternatively, we may be more concerned wit general rules of choice, rather than with the solution of a specif problem. We have, for example, the rule that, under certain cond tions, it is better to order fixed amounts at irregular intervals of tim (provided of course that this is possible in practice) rather tha irregular amounts at fixed intervals.

As a rule studies of the kind referred to above have been large concerned with the operating characteristics of specific kinds of stor Another important aspect, especially from the architectural and finan cial points of view, is the problem of making stores as small as co venience will allow. This was first brought to my attention by th Nuffield Foundation's Division for Architectural Studies, chiefly connection with the design of scientific laboratories. In principle th standard mathematical theory should be capable of application her The main difficulties are the complexity of the standard theory, an the fact that, in the context of scientific research at any rate, it is ve difficult to assign suitable cost-functions. What seems to be urgent

quired is a fairly simple approach that could be used by those without
tensive mathematical knowledge.

The problem of overall store-size is quite a formidable one for any
chitect who has, for instance, to design a group of laboratory build-
gs. We are not, of course, concerned here with individual details of
sign. The essence of the matter is to know how much space should
allotted to a general store serving a number of buildings. Similar,
ough less pressing, difficulties also arise in connection with localised
orage space in each laboratory unit. When a new group of buildings
to be erected one would like to know, first of all, what the pattern of
mand for the various materials used is likely to be. This in itself is a
ry difficult question, and the necessary vagueness of any estimates of
obable demand is sufficient to make a highly sophisticated mathe-
atical treatment hardly worth while.

It might seem from these remarks that, in the absence of a precise
rmulation of the problem, the question 'How big should the store be?'
ight just as well be left to architectural intuition. This, however, is a
unsel of despair, and will not, in any case, prevent architects asking
atisticians the awkward question and expecting some kind of con-
uctive answer. Decisions have to be made, whatever happens, and
is certainly worth looking for a method of analysis which might shed
certain amount of light on the problem.

In the first place, certain standards of storage accommodation are
ready current. One might, therefore, hope to discover whether
isting stores are adequate or whether they could be made smaller
thout serious inconvenience. This would give some general guidance
to the soundness of present practice. Such results could be applied
st in a broad way to the design of new stores. Then, if more precise
formation were available about the probable pattern of demand, one
ight hope for a more exact criterion.

It is the purpose of the present paper to suggest a simple way of
proaching these problems. First, a rough deterministic formulation
the process in question will be given. This is useful in helping to
rm a preliminary picture of the situation, but it should not be relied
on as a basis for accurate calculation unless very large numbers are
volved. Then an elementary probability model will be introduced
nich allows us to relate a performance index or level of satisfaction
stores of different sizes. To begin with, we envisage a store containing
ly one kind of material. Subsequently, the approach is extended to
ver the storage of several different kinds of material.

I have not yet had the opportunity to undertake the required surveys
current practice and to try out the methods recommended, but the
tter would appear to have possible applications over quite a wide
ld. It is hoped, therefore, that this paper will encourage those who
e concerned with storage problems to discover how well these
inciples can be applied in practice, and to see whether any major
adequacies can be remedied by simple modifications.

Deterministic Model

Let us first consider the situation where the amount of materia stored can be regarded as a continuous variable. (This would b strictly true with liquids that could be supplied in any quantity re quired, and not merely in containers of fixed size; and might b approximately valid when fairly large numbers of separate objects wer involved.) We shall also suppose that the demand is steady and con tinuous. This means that we can use a simple deterministic formulation The assumptions are in general unrealistic, but will serve for a pre liminary examination of the problem. Moreover, the model might b useful in practice if the numbers involved were sufficiently large fo random fluctuations to be negligible.

Now if the demand is steady and continuous, it is predictable wit considerable accuracy. We can, therefore, arrange for regular order to be placed (in advance if necessary) so that as soon as the number o items in the store has fallen to some predetermined value, a, it is imme diately replenished to a fixed maximum, b. A graph of the actua contents of the store at any time thus has the saw-tooth form shown i Fig. 1. If we choose the time-scale so that the intervals between order are all one unit, the demand per unit of time is $d = b - a$.

We see immediately from Fig. 1 that the store-size is much too large The store never contains less than the quantity a, and the holding o these stocks under the conditions specified offers no advantage what ever. It is clear that, apart perhaps from a small safety margin, the ful capacity need be no more than d. Replenishment could then occu when the total stock had fallen to zero. This state of affairs is illustrate in Fig. 2. A considerable saving in storage capacity would result if were large, without, moreover, any adverse effect on the ability t meet the demand.

Further saving will be achieved if we can reduce the time interva between successive orders so as to make the 'saw-teeth' smaller. If w could order three times as frequently as in Fig. 2 we should obtain th kind of stock-size pattern appearing in Fig. 3.

There is, of course, a limit to such contraction. First, orders canno be placed more closely than the delivery routine of the wholesaler wi allow. This lower limit may be a matter of days, weeks, or months Secondly, as soon as we get somewhere near the state of affairs repre sented by Fig. 3 (or even Fig. 2, if d is small) random fluctuations i demand will begin to be important. A deterministic model will the no longer give an adequate representation of the changes in stock-siz with time.

Nevertheless, the above discussion does suggest the following tw tentative rules for reducing store-size as much as possible.

(*a*) For a given interval between the arrival of successive orders, th store-size need only be sufficient just to prevent the stock runnin out.

) It is, in general, advantageous to have frequent small orders, rather than infrequent large orders.

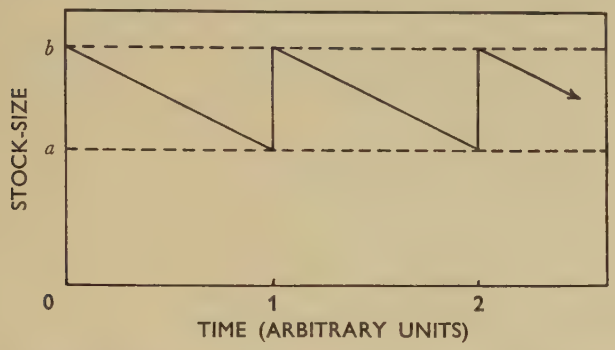
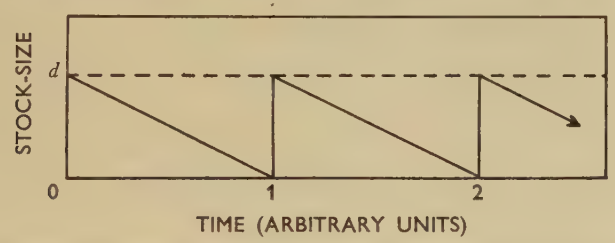
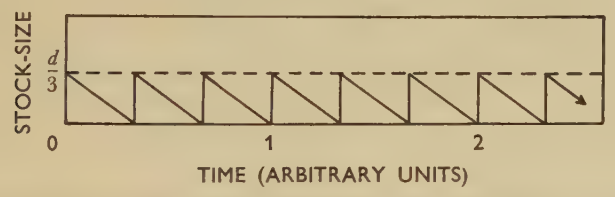


Fig. 1. Fluctuation of stock-size with time; deterministic model with large store-size.



g. 2. Fluctuation of stock-size with time; deterministic model with store-size allowing no margin (time-scale as for Fig. 1).



g. 3. Fluctuation of stock-size with time; deterministic model with time-scale as Fig. 2 and more frequent replenishment.

These rules are really common sense, but they can be applied only o to a certain point, depending on the limitations already mentioned ove. The deterministic model we have been discussing is useful for rming an initial picture of the process in question. But as soon as ctuations in demand begin to make themselves felt we must adopt a obability model. This will entail a new standard of judgement, since ere will always be the possibility of the stock running out before a new ler arrives. The problem is to make the chance of this happening ficiently small.

obability Model

It is not proposed to investigate here any very elaborate probability odel. We shall use only a simple formulation to show the sort of alysis and interpretation involved in this way of looking at the

problem. More sophisticated developments might be tried later if they seem to be required.

We shall now drop the assumption of a continuously variable quantity of material in the store and suppose that discrete items are involved. The maximum capacity of the store is taken to be s units. But we now postulate a random pattern of demand such that requests for individual items from the store follow a Poisson process for which the average number of requests per unit of time is d.

A salient feature of the model is the kind of ordering policy to be adopted. We want something simple but not hopelessly unrealistic. Moreover, we also have to decide what to do about requests which arrive when the store is empty. One way of dealing with the latter is to allow these requests to accumulate on a kind of waiting-list which is identified with a *negative* value of stock-size. If, with this assumption, we envisage regular orders of fixed amounts, it is easily seen that fluctuations in stock-size will increase without limit as time goes on. This is unsatisfactory, as most storage systems are to some extent self-regulating. Another possibility would be to have a fresh order of amount s arrive whenever the stock-size drops to zero. This also seems unsatisfactory. Not only is it too unrealistic, but it also entails several successive orders sometimes arriving at very short intervals.

Some compromise is required, and the one chosen for investigation is as follows. Orders are placed at regular intervals, but are of variable size, being in fact just sufficient to bring the store up to capacity. In other words, we envisage the dates of deliveries from the wholesaler as fixed in advance, but suppose that the actual size of the order required can be determined at the last minute and rapidly communicated to the wholesaler, say by telephone. Fluctuations in stock-size now follow a pattern like that shown in Fig. 4. In the first period of time the stock is not exhausted before being replenished. But in the second period negative values of stock-size occur, which we have agreed to interpret as a waiting-list of demands.

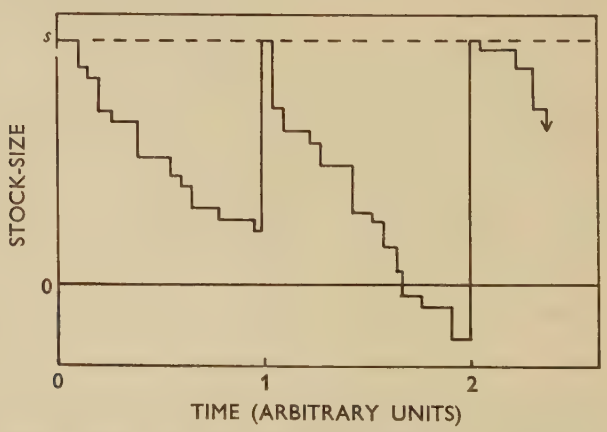


FIG. 4. Fluctuation of stock-size with time; probability model with regular replenishment to full capacity of store.

The advantage of this model is that it involves statistical fluctuation thout being over-elaborate and mathematically intractable. Although e ordering policy is a trifle artificial, it at least has the merits of oiding instability, i.e. indefinite increase in variance of stock-size, d of permitting investigation of the extent to which some demands nnot be met immediately if the size of the store is small.

In general, for a given fixed interval between successive orders, we ant the store-size, s, to be as small as possible consonant with some propriate criterion of efficiency. We could, for example, try to keep a minimum the proportion of intervals like the second one in Fig. 4. is, however, is unnecessarily stringent. Even if all intervals were like is, but the stock ran out only near the end of the period, compara-ely little inconvenience would be caused. A better index of per-rmance is the efficiency, E, defined as the proportion of requests that n be met immediately. This is the criterion we shall adopt. Its ethod of calculation is indicated below, and a convenient table is ovided for quick reference. One might do better still by working ith the average amount of waiting involved, or by using some other pect of the distribution of waiting-time. Such a procedure leads to mewhat tedious computations, and it is by no means certain that it rries any substantial advantages.

athematical Theory

The pattern of demand is, as stated above, taken to follow a Poisson ocess. If the average number of requests per unit of time is d, and if e choose the time-scale so that the interval between successive orders one unit, then the probability of exactly r requests in an interval is ven by the Poisson frequency distribution

$$\mathrm{p}(r,d) = \frac{\mathrm{e}^{-d}d^r}{r!},\ 0 \leqslant r < \infty \qquad \ldots.(1)$$

e shall also need the accumulated frequencies defined by the nctions

$$\mathrm{P}(s,d) = \sum_{r=s}^{\infty} \mathrm{p}(r,d) \qquad \ldots.(2)$$

th of the quantities in (1) and (2) have been tabulated by Molina.[4] Now, if in any interval $r < s$, all r requests can be met immediately. on the other hand, $r \geqslant s$, only s requests can be dealt with straight ay. The average number of requests per interval receiving imme-ate attention is therefore

$$\mathcal{N} = \sum_{r=0}^{s-1} r\mathrm{p}(r,d) + \sum_{r=s}^{\infty} s\mathrm{p}(r,d) \qquad \ldots.(3)$$

nce the average overall demand per interval is d, the efficiency hieved is

$$E = \mathcal{N}/d \qquad \ldots.(4)$$

We now substitute (3) in (4) and make use of (1) and (2). After little rearrangement we find that we can write

$$1 - E = \mathrm{p}([s-1],d) - \left(\frac{s}{d} - 1\right)\mathrm{P}(s,d) \qquad \ldots\ldots(5)$$

where $1 - E$ is the probability that a request will not be satisfie immediately. Evaluation of these probabilities is quite easy if we us (5) in conjunction with Molina's Tables.[4]

In practical applications we want to know, for a given value of th demand per interval, what store-size is required to achieve a specifi efficiency. Thus suppose we know that d is about 12 and want t arrange that not more than one request in a thousand fails to be satis fied immediately. Direct solution of (5) is not possible, but we coul easily experiment with a few trial values in the usual way, interpolatin inversely between probabilities straddling 0·001. We actually find tha s is then approximately 21. It seemed worth while to tabulate som results of this kind. Values of $1 - E$ were first calculated for severa adjacent integral values of s when $d = 1(1)20$. As $1 - E$ falls o roughly exponentially, inverse logarithmic interpolation was used t find the values of s, given d, corresponding to $1 - E = 1/20$, $1/100$ 1/1 000 and 1/10 000. These results are shown in Table I. Simila calculations were also performed for the range $d = 5(5)100$, usin s-values in multiples of 5. Table II gives the final figures.

TABLE I

This gives the store-size, s, corresponding to a given value of the demand, d, per interval and chosen level of efficiency, E, for $d = 1(1)20$

Average Demand per Interval d	Proportion of Requests not met Immediately $(1 - E)$			
	1/20	1/100	1/1 000	1/10 000
1	2·5	3·5	4·8	6·0
2	3·7	5·1	6·8	8·2
3	4·9	6·5	8·5	10·2
4	6·0	7·8	10·0	12·0
5	7·0	9·1	11·5	13·6
6	8·1	10·3	13·0	15·2
7	9·1	11·5	14·3	16·8
8	10·1	12·7	15·7	18·3
9	11·1	13·9	17·1	19·8
10	12·1	15·0	18·4	21·2
11	13·1	16·2	19·7	22·6
12	14·1	17·2	21·0	24·0
13	15·1	18·4	22·3	25·4
14	16·1	19·6	23·5	26·7
15	17·1	20·7	24·8	28·1
16	18·0	21·8	26·0	29·4
17	19·0	22·9	27·2	30·8
18	20·0	24·0	28·4	32·1
19	21·0	25·1	29·7	33·4
20	21·9	26·2	30·9	34·7

If we use E as the criterion of efficiency, the results obtained above ill be valid for a slightly more general model than the one specified. t will not be necessary for the demand to be a strict Poisson process n continuous time: it will be sufficient if the total numbers of requests n each interval between successive orders follow a Poisson distribution.

TABLE II

his gives the store-size, s, corresponding to a given value of the demand, d, per interval and a chosen level of efficiency, E, for $d = 5(5)100$

Average Demand per Interval d	Proportion of Requests not met Immediately $(1 - E)$			
	1/20	1/100	1/1 000	1/10 000
5	7	9	12	14
10	12	15	18	21
15	17	21	25	28
20	22	26	31	35
25	27	31	37	41
30	31	37	43	47
35	36	42	48	53
40	41	47	54	59
45	46	53	60	65
50	51	58	65	71
55	55	63	71	77
60	60	68	76	83
65	65	73	82	89
70	70	78	87	95
75	74	84	93	100
80	79	89	98	106
85	84	94	104	112
90	88	99	109	117
95	93	104	115	123
100	98	109	120	128

iscussion

First of all it should be noted that we are mainly concerned with uations in which the average demand per interval is fairly small. his is because statistical variation is small with a sufficiently large erage demand and prediction is comparatively accurate: deter-inistic calculations can then be applied as already mentioned above, t it is uncertain how often this possibility may occur in practice.

Suppose for a particular kind of object the wholesaler can make liveries once a week, but not more frequently. We start by taking s as the basis for repeated orders. If we are investigating an existing ilding, we next examine the store records. Suppose the average mand per week is found to be about 6 units. Reference to Table I ows that, with a store-size of 8, only one request in 20 would fail to met immediately. The choice of efficiency level is, of course, some-at arbitrary, and we may prefer to work with levels of 1 in 100 or n 1 000 to be on the safe side. The latter would be achieved by a

store-size of 13, i.e. little more than double the average demand be tween successive orders. If the average demand per interval is highe relatively less stock must be held, and conversely. With non-integra values of the average demand, rough interpolation in the tables sufficient to obtain the required results.

It is important that the average demand should be known wit reasonable accuracy. Assuming that the demand is Poisson in charac ter, we can easily determine confidence limits for our estimate of th average and can refer these values to the tables. In general we sha probably want to use a sample large enough to reduce the coefficient variation of the average demand to about ten per cent or less. Some times the average demand will not be steady, but will have some kin of systematic fluctuations like seasonal variations. This raises problem that are beyond the scope of the present discussion.

The probability model we have adopted supposes that the materia with which we are concerned is stored and supplied in discrete unit It is important in applications to make sure that we use the prope units. Thus in dealing with test-tubes the appropriate working un will normally be a whole box rather than a single test-tube. Similarly many acids are supplied in containers of standard size.

The theory described above applies immediately only to one part cular kind of material. A great many stores, however, are of a genera purpose kind. The most obvious way of handling this problem is t start by treating each kind of object separately. As this might result i a prohibitively large number of separate items, the best plan is to tr to divide the stock into a manageable number of homogeneous group We then examine each group separately. In considering any particula group we start by applying the two rules suggested in the section o deterministic models. If the average demand for the smallest feasibl interval between two successive orders is large, the elementary dete ministic account may be quite adequate for calculating the smalle workable store-size for the group in question. But if the averag demand per interval is small we must base the store-size on a probabilit model. Having fixed an acceptable chance of failure to meet deman we refer to Table I or II as before. The sum of the store-sizes for th individual groups then gives the storage capacity required for the whol building.

This approach tacitly assumes that all the different kinds of objec in a single, supposedly homogeneous, group can be regarded as inte changeable so far as storage space is concerned. To some extent th space not required at any time by one material may be available fo another, but there are restrictions depending on the nature of th materials, store layout, type of containers, safety precautions, et It is clear that any material requiring specialised treatment must b considered separately.

It is at present unknown what the likely consequences of the abov procedure would be. One gains a general impression that many stor

ld stocks vastly in excess of a reasonable capacity, and that drastic ductions could be made without materially affecting the service pro-ded. Storage space is often very expensive and large stocks are liable deteriorate if they are replaced only very slowly. There would seem, erefore, to be considerable economic advantage in attempting to ply systematically the kind of approach recommended in this paper. here is, of course, the point that a small store with rapid turnover and tailed running checks on stock held is more difficult to organise cause of the increased clerical work involved in having a more borate ordering policy. Nevertheless, it may be possible to find ways streamlining the work.

The foregoing discussion has been mainly related to the problem of ecking the performance of existing stores. A similar kind of situation ises when we are concerned with the store for a new group of build-gs, except that we then have the added difficulty of trying to predict e approximate demand for each of a number of groups of items. The sic principles, however, remain the same.

knowledgement

I am greatly indebted to members of the Nuffield Foundation's vision for Architectural Studies for drawing my attention to these oblems and for many helpful discussions.

REFERENCES

ACKOFF, R. L. (1956). 'The development of Operations Research as a science', *Operations Research*, **4**, 265.

GANI, J. (1957). 'Problems in the probability theory of storage systems', *J. R. Statist. Soc.*, B, **19**, 181.

KENDALL, D. G. (1957). 'Some problems in the theory of dams', *J. R. Statist. Soc.*, B, **19**, 207.

MOLINA, E. C. (1942). *Poisson's Exponential Binomial Limit*. Van Nostrand, New York.

FACTOR ANALYSIS

GEOFFREY H. JOWETT
University of Sheffield

In this article Dr Jowett gives a simple description of th essentials of factor analysis, with particular reference to th centroid method. Despite the fact that some statisticians di approve of the subject, we feel that there is ample justificati for an expository article of this kind, so that users of statisti may be encouraged to discover whether this method of analys is helpful in any of their problems.

Factor analysis has long been regarded as the black sheep of th statistical fold, partly because of the crudity of some of its mathe matical theory, and partly because of the indeterminacy, from strictly objective point of view, of the conclusions which can be draw from it. In spite of this, psychologists have made valuable use of it i comparing and interpreting aptitude tests; also its wider applicabilit has been indicated in a recent article in this journal[2] where Dr H Harper has described its use for interpreting the mutual relationshi between fourteen tests of physical properties of Cheshire cheeses. Y another potential application is to the joint interpretation of physic and chemical tests on coke.

This paper is an attempt to clarify the basic essentials of the tech nique, showing how to use it in simple cases, why it works and what i results mean. There seems to be a real need for a short exposition this kind; most standard textbooks on elementary statistics boycott th subject, or are vague about the basic assumptions, whereas such mono graphs as do go properly into the theoretical background are apt to b mathematically indigestible, or written primarily for psychologists. hope that a few more applied statisticians will be encouraged to con sider whether or not factor analysis is relevant to any of their problem and to give it a critical trial in one or two simple cases.

Factor analysis is essentially a method of comparing and combinin the results of a number of different tests, all carried out on the sam large group of individuals. We meet it in a rudimentary form whe school marks are averaged to give a measure of general academ ability, and the sophistication begins to creep in when the woodwo mark is given a weighting of only, say, a half in this averaging on th ground that it is not so academic as the other subjects. Since attainme tests lie within the experience of most people, and also since the formed the main stimulus for the development of the subject, it convenient to assume, in an introductory exposition of factor analysi that the data to be analysed consist of the scores in several attainme

sts of a large group of candidates representative of some well-fined population, such as schoolboys ready to sit for the 'eleven plus' amination.

alysis involving One Factor

I shall denote the mark scored by candidate (q) in test p by the mbol $x_{p(q)}$, and shall assume that the marks for each test have been ndardised (i.e. expressed as deviations from the overall mean for the st in units of the standard deviation); thus the standardised marks for ch test will have zero mean and unit standard deviation. Let us etend for the moment that some fundamental ability, such as the telligence' of each candidate, is capable of direct measurement, and n be given a standardised score $i_{(q)}$ for candidate (q). If the regression x_p on i is linear, we may write

$$x_{p(q)} = \alpha_p i_{(q)} + \epsilon_{p(q)} \qquad \ldots.(1)$$

ere α_p is the coefficient of regression of x_p on i in the population of ndidates (i.e. the 'effect' of intelligence on the score in that particular t), and $\epsilon_{p(q)}$ is that residual part of the score which is independent and unpredictable from, intelligence. In general it is likely that hough all the sets of residuals are uncorrelated with i they will not be uncorrelated with each other, perhaps because some tests semble others, having other things in common besides the tendency measure intelligence; but for the moment we shall assume that they e diverse enough in character for any such correlations to be negli-ble. Thus $i_{(q)}$, $\epsilon_{1(q)}$, . . ., $\epsilon_{k(q)}$ will be taken to vary independently one another in the population of candidates.

Now is the time to drop our pretence. In practice we cannot measure directly. Yet we can conceive of such a measure, and of a set of t variables related to it and to one another as described in the last ragraph. Moreover, under these assumptions of independence we n obtain estimates of the regression coefficients α_p (called the *loading* the *factor* i in the test p) by the method of factor analysis, and these able us to estimate a value for $i_{(q)}$ for any particular candidate. Thus we divide $x_{p(q)}$ by α_p we have an estimate of $i_{(q)}$—here $\epsilon_{p(q)}$ plays e role of an error of estimation—and by averaging all such estimates m the k tests we obtain a more trustworthy estimate of $i_{(q)}$. The erage would, of course, usually be a weighted average and will be scribed further below. The whole procedure is reminiscent of eraging school marks to assess academic attainment, but on a much ner plane of sophistication. At first glance it seems as if we are gically obtaining an estimate of an undefined quantity, but this is illusion; the statistical nature of the factor, though not its psycho-ical meaning, is made perfectly definite by our assumptions about ependence.

The estimation of the factor loadings α_p forms the fundamental blem of factor analysis, and will now be considered. It is easy to

show, under the independence conditions which have been stated, tha the population correlation coefficient $\rho_{pp'}$ between test variables x and $x_{p'}$ is given by the equation

$$\rho_{pp'} = \alpha_p \alpha_{p'} \qquad \ldots . (2$$

and therefore that for the matrix of all possible intercorrelation between the k tests,

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \cdots & \rho_{1k} \\ \rho_{21} & 1 & \rho_{23} \cdots & \rho_{2k} \\ . & . & . \ . \ . & . \\ . & . & . \ . \ . & . \\ . & . & . \ . \ . & . \\ \rho_{k1} & \rho_{k2} & \rho_{k3} \cdots & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & \alpha_1\alpha_2 & \alpha_1\alpha_3 \cdots & \alpha_1\alpha_k \\ \alpha_2\alpha_1 & 1 & \alpha_2\alpha_3 \cdots & \alpha_2\alpha_k \\ . & . & . \ . \ . & . \\ . & . & . \ . \ . & . \\ . & . & . \ . \ . & . \\ \alpha_k\alpha_1 & \alpha_k\alpha_2 & \alpha_k\alpha_3 \cdots & 1 \end{pmatrix} \ldots . (3$$

There are only k unknown α's, whereas there are $\frac{1}{2}k(k-1)$ ρ's, so tha if k is three or more the α's are determinate and could be calculated the ρ's were given. In practice, however, we have only a sample fro the infinite population of potential candidates, and the correlations a subject to sampling errors. Hence the α's cannot be chosen so as secure exact agreement with all the ρ's and some kind of averagi method has to be used, in the same way as the method of least squar is used to obtain the 'best' solution of a set of simultaneous equatio with observed coefficients when there are more equations than u knowns. The *centroid* method of Professor Thurstone[3] is an easy way doing this which has certain optimal properties, and is applicable if t factor loadings are predominantly positive; there exist other metho preferable on mathematical grounds, but this one is especially suitab for beginners.

Suppose that in place of the diagonal 1's on the two sides of (3) had $\alpha_1^2, \alpha_2^2, \ldots, \alpha_k^2$; then by equating the average entries in the fi columns, for example, we should have

$$(\alpha_1^2 + \rho_{21} + \rho_{31} + \ldots + \rho_{k1})/k = \alpha_1 \bar{\alpha} \qquad \ldots . ($$

where $\bar{\alpha}$ is the mean of the k factor loadings.

Averaging all k such equations, we have

$$\left\{ \begin{array}{l} \text{Average of column averages} \\ \text{on the left-hand side of (3)} \\ \text{with } \alpha^2\text{'s in diagonal} \end{array} \right\} = \bar{\alpha}^2 \qquad \ldots . (4$$

Thus α_1 is the first column average divided by the square root of t average for the whole table given by (5); the rest of the α's are dete mined in the same way. This is a circular procedure, because we not know the α's to start with; but if we can make an initial guess at a particular α^2, any error we commit is divided by k in the colu average and by k^2 in the table average, so that our deduced value m be expected to be a better approximation to the correct value. what is done is to guess the α^2 values (the highest correlation in ea

umn is a good first guess), use (4) and (5) to get a better approxi-tion, and then to iterate until two successive approximations agree, shown in Table I, where the process is demonstrated on the corre-ons between five test variables.

TABLE I

Single-factor analysis for five test variables

Test \ Test	Correlations 1	2	3	4	5	Total	$\frac{1}{\sqrt{(\text{Total})}}$
1	—	0·51	0·31	0·60	0·44		
2	0·51	—	0·38	0·76	0·59		
3	0·31	0·38	—	0·39	0·35		
4	0·60	0·76	0·39	—	0·71		
5	0·44	0·59	0·35	0·71	—		
iessed α_p^2 ..	0·60	0·76	0·39	0·76	0·71		
otal	2·46	3·00	1·82	3·22	2·80	13·30	0·2742
t. α_p = 1st approx.	0·67*	0·82	0·50	0·88	0·76		
vised α_p^2	0·45	0·67	0·25	0·77	0·58		
vised total ..	2·31†	2·91	1·68	3·23	2·67	12·80	0·2795
t. α_p = 2nd approx.	0·64	0·81	0·47	0·90	0·75		
vised α_p^2	0·41	0·66	0·22	0·81	0·56		
vised total ..	2·27	2·90	1·65	3·27	2·65	12·74	0·2802
t. α_p = 3rd approx.	0·64	0·81	0·46	0·92	0·74		
vised α_p^2	0·41	0·66	0·21	0·85	0·55		
vised total ..	2·27	2·90	1·64	3·31	2·64	12·76	0·2800
t. α_p (final) ..	0·64	0·81	0·46	0·93	0·74		

* 0·67 = 2·46 × 0·2742, etc.

† 2·31 = 2·46 − 0·60 + 0·45, etc.

Test \ Test	Residuals 1	2	3	4	5
1	—	−0·01	0·02	0·00	−0·03
2	−0·01	—	0·01	0·01	−0·01
3	0·02	0·01	—	−0·04	0·01
4	0·00	0·01	−0·04*	—	0·02
5	−0·03	−0·01	0·01	0·02	—

* −0·04 = 0·39 − 0·46 × 0·93.

n estimate of the value of $i_{(q)}$ for any candidate is then obtained by averaging procedure indicated above. Each test score is divided by

the appropriate factor loading, and weighted (if we want the bes result) with the reciprocal of the estimated variance of ϵ_p/α_p. Thus

$$\text{Est}[i_{(q)}] = \frac{\frac{\alpha_1^2}{\text{var } \epsilon_1} \cdot \frac{x_{1(q)}}{\alpha_1} + \frac{\alpha_2^2}{\text{var } \epsilon_2} \cdot \frac{x_{2(q)}}{\alpha_2} + \ldots + \frac{\alpha_k^2}{\text{var } \epsilon_k} \cdot \frac{x_{k(q)}}{\alpha_k}}{\frac{\alpha_1^2}{\text{var } \epsilon_1} + \ldots + \frac{\alpha_k^2}{\text{var } \epsilon_k}} \quad \ldots . (6)$$

$$= \frac{\frac{\alpha_1}{1-\alpha_1^2} x_{1(q)} + \frac{\alpha_2}{1-\alpha_2^2} x_{2(q)} + \ldots + \frac{\alpha_k}{1-\alpha_k^2} x_{k(q)}}{\frac{\alpha_1}{1-\alpha_1^2} + \ldots + \frac{\alpha_k}{1-\alpha_k^2}} \quad \ldots . (7)$$

The values α_p^2, which may be interpreted as the proportions of the variation of each variable caused by the factor i, called the *common factor* of the tests, are called *communalities*. So far it has been assumed that i is to be interpreted as intelligence, but its meaning will clearly depend on the nature of the battery of tests; if they were all arithmetic tests, the common factor would have the narrower interpretation of 'arithmetical ability'. It should never be forgotten that the common factor of a group of tests determined in this way can never be more than a kind of average of test scores; the different weighting coefficients merely result in more weight being given to the tests which, as it were, toe the party line, and less to deviationists.

Since the centroid method involves iteration, square roots, and the averaging of product-moment correlations, it goes almost without saying that sampling problems, like the determination of standard errors of factor loadings and tests of goodness of fit, are difficult, and at the elementary stage are usually passed over rather lightly; for the factor analysis to be taken as trustworthy the sample of candidates has to be large. Moreover, the population sampled must have a clear and useful specification, since the value of the factor for an individual is because of standardisation, merely relative.

When the factor loadings have been estimated, the *residual correlation* between any pair of test variables may be calculated by subtracting from the original correlations the product of the corresponding pair of factor loadings, as shown at the foot of Table I. This is the correlation between the residual variables ϵ_p, and under our assumptions it should be zero except in so far as it is affected by the sampling errors. If all the residuals are small (i.e. comparable with the standard errors of the original correlations) and show no pattern, we usually take it that sampling errors are the correct explanation.

Analysis involving Two Factors

In Table II is shown a factor analysis like that of Table I, carried out

n a matrix of correlations between six tests, for which the intercorrela-ions are given by Sir Cyril Burt in his book *The Factors of the Mind.*[1] The methods of analysis used by Burt are somewhat different from nose given here, though the essentials are the same.) The residual orrelations have a clear pattern; those between 'verbal' tests 1, 2, nd 3 are positive and those between 'practical' tests 4, 5, and 6 are lso positive, whereas correlations between pairs of tests taken one from ach group are negative. This suggests that the single-factor hypothesis inadequate.

TABLE II

Single-factor analysis for six tests: Burt's data

Test	Correlations					
Test	Composition	Reading	Spelling	Handwork	Drawing	Writing
1. Composition ..	—	0·58	0·49	0·30	0·38	0·44
2. Reading ..	0·58	—	0·46	0·10	0·13	0·21
3. Spelling ..	0·49	0·46	—	0·09	0·12	0·25
4. Handwork ..	0·30	0·10	0·09	—	0·50	0·28
5. Drawing ..	0·38	0·13	0·12	0·50	—	0·36
6. Writing ..	0·44	0·21	0·25	0·28	0·36	—
Est. α_p	0·87	0·52	0·49	0·43	0·52	0·55
	Residuals					
1. Composition ..	—	+0·13	+0·06	−0·07	−0·07	−0·04
2. Reading ..	+0·13	—	+0·21	−0·12	−0·14	−0·08
3. Spelling ..	+0·06	+0·21	—	−0·12	−0·13	−0·02
4. Handwork ..	−0·07	−0·12	−0·12	—	+0·28	+0·04
5. Drawing ..	−0·07	−0·14	−0·13	+0·28	—	+0·07
6. Writing ..	−0·04	−0·08	−0·02	+0·04	+0·07	—

The situation is in many ways analogous to that encountered in olynomial regression; there the simplest analysis is the fitting of a raight regression line, but if points on the left and right of the scatter agram tend, for example, to fall above the line, while points in the iddle tend to fall below, it is inferred that the straight line is inade-uate to describe the data, and a second-degree term must be fitted. factor analysis, when our residuals show a pattern, we postulate the istence of a second factor. Moreover, just as a straight regression line ll give a crude approximation to a curvilinear relationship, so a gle factor will give a crude approximation to the structure of a ttery of tests, even when the single-factor hypothesis is wrong, and centroid method, like least squares in regression, ensures that this ude explanation shall be the best of its type. The role of the second-gree orthogonal polynomial in regression is paralleled in factor alysis by a second factor, of which the loadings are obtained by the plication of the centroid method a second time, in slightly modified m, to the matrix of residuals.

The centroid method cannot be applied directly, since the loadin in the second factor are not predominantly positive, and the positi and negative residuals would tend to cancel out in the column sum If, however, any variable likely to have negative loadings in the secon factor has its sign changed (is *reflected*, in other words), the elements the matrix in the same row and column will have their signs chang too. It is recommended that this reflecting should be carried out such a way as to secure marginal totals as great as possible; then th second factor, which is uncorrelated with the first, will explain aw the maximum amount of residual correlation. For the residuals Table II it is clearly best to reflect test variables 4, 5, and 6. T squared factor loadings are then guessed as before, and iterated to fin estimates as in Table I. When final estimates are secured, as shown Table III, their signs are reflected back again when appropriate. T

TABLE III

Analysis for second factor: Burt's data

Test \ Test	Residuals from First Factor (Reflected)					
	Composition	Reading	Spelling	(−) Handwork	(−) Drawing	(−) Writin
1. Composition ..	—	0·13	0·06	0·07	0·07	0·04
2. Reading ..	0·13	—	0·21	0·12	0·14	0·08
3. Spelling ..	0·06	0·21	—	0·12	0·13	0·02
4. (−) Handwork	0·07	0·12	0·12	—	0·28	0·04
5. (−) Drawing ..	0·07	0·14	0·13	0·28	—	0·07
6. (−) Writing ..	0·04	0·08	0·02	0·04	0·07	—
Est. loadings ..	0·21	0·44	0·33	0·40	0·45	0·14
	Composition	Reading	Spelling	Handwork	Drawing	Writin
Est. β_p	0·21	0·44	0·33	−0·40	−0·45	−0·14
	Residuals from Two Factors (with Correct Sign)					
1. Composition ..	—	+0·04	−0·01	+0·01	+0·02	−0·01
2. Reading ..	+0·04	—	+0·06	+0·06	+0·06	−0·02
3. Spelling ..	−0·01	+0·06	—	+0·01	+0·02	+0·03
4. Handwork ..	+0·01	+0·06	+0·01	—	+0·10	−0·02
5. Drawing ..	+0·02	+0·06	+0·02	+0·10	—	+0·01
6. Writing ..	−0·01	−0·02	+0·03	−0·02	+0·01	—

communality of any test, i.e. that portion of its variance which is d to the two factors, is now given by $\alpha_p^2 + \beta_p^2$, where β_p is the loadi of the test in the second factor. New residuals may be calculated subtracting products of second factor loadings from the residual correl tions, just as in the case of the first factor; these are shown at the botto of Table III; if these are comparable with the sampling errors of t original correlations, and devoid of pattern, factorisation may adjudged complete.

Estimation of the joint values of the two factors may be carried out as lows. Writing $u_{(q)}$ as the second factor, we have

$$x_{p(q)} = \alpha_p i_{(q)} + \beta_p u_{(q)} + \eta_{p(q)}, \text{ say,} \quad \ldots.(8)$$

here $\eta_{p(q)}$ is now the residual part of $x_{p(q)}$ unexplained by either ctor, and has variance $1 - \alpha_p^2 - \beta_p^2$. Dividing through by this antity, we obtain a new relation of the form

$$x_p'_{(q)} = \alpha_p' i_{(q)} + \beta_p' u_{(q)} + \eta_p'_{(q)} \quad \ldots.(9)$$

here $\eta_p'_{(q)}$ now has unit variance, and obtain estimates of $i_{(q)}$, $u_{(q)}$ by ultiple regression, taking the k numbers $x_p'_{(q)}$ (fixed q) as values of e dependent variable, and the two sets of reduced loadings α_p', β_p' two independent variables, which will usually be almost uncorre-ted. Uncorrected sums of squares and products are used, there being constant term in (9), and we note in passing that formula (7) could ually well have been derived by the corresponding simple regression. When factor analysis involves more than one factor, however, our us of interest changes. We become less interested in evaluating the tors for the individual candidate, and more interested in the relations tween the tests themselves from the point of view of factor composi-n, a trivial problem when there is only one factor. There are useful plications of this: for example, we may start off with a large experi-ntal battery of tests, and aim at choosing from them the best skeletal for use in routine work (e.g. personnel selection) later; or we may sh to identify the combination of attributes measured by certain ecified tests in the battery.

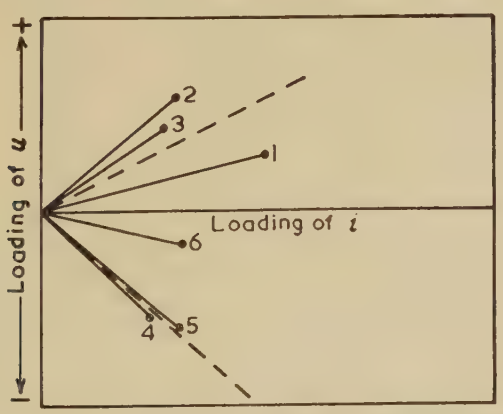


FIG. 1. Burt's example: scatter diagram for factor loadings in the factors i and u.

We can obtain a very informative picture of the mutual relationships tween the tests by plotting the loadings of the two factors in each test ainst one another on a scatter diagram, as in Fig. 1, which illustrates rt's example; there is one point for each test, and it is interesting to te that the part of the correlation between two tests which may be

ascribed to the two factors is equal to the product of the distances of t two test points from the origin and the cosine of the angle betwe them. If we were satisfied that for our prognostic purposes all t information required was given by the two factors, we could b economise on tests, with some loss of accuracy of course, by throwi out tests 3, 4, and 6. Tests 2 and 5 would be retained rather than te 3 and 4 respectively since they have approximately the same fact composition but greater communalities (i.e. same direction but grea distance from the origin); test 1 would be retained because of its hi communality and test 6 rejected for the opposite reason. From t point of view of identification of attributes contained in the tests note that tests 1, 2, and 3 are predominantly in the same directio suggesting a common basis which could be described as 'literary'; a tests 4 and 5 are close to a second direction, suggesting a common ba which could be described as 'handicraft'. Evidently test 6 partak more or less equally of both bases (as might be expected from description), though rather weakly.

Changing to Interpretable Factors

From the point of view of identification of the factors with son nameable ability, it might have been more satisfactory if the two pr dominant directions just described could have coincided with the t axes, instead of being more nearly internal and external bisectors of t angle between them. All the tests except test 6 would then have h strong loadings in one factor and weak loadings in the other, and t factors could have been identified with literary and handicraft abiliti As things stand, the very nature of the centroid method has forced t first factor *i* to be a mongrel affair; it lies, as it were, roughly two-fift of the way between literacy and handiness, and its position could pushed still further towards the literary by including one or two ext literary tests (e.g. comprehension of literary passages) in the batter Like the slope of the straight regression line used as an approxim tion to a parabolic regression, which could be changed by adding da at one end of the range of the independent variable, it has little meani in itself, and is only a means to an end, helping to supply a co-ordina system with the aid of which the tests may be plotted, compared, a their nature recognised. The situation is in direct contrast to that e countered in a single-factor battery of tests, where the single factor at once interpretable as the common basic ability required to obta a good score in a group of otherwise independent tests.

This lack of practical significance of the factors taken individually not always a matter for concern in factor analysis, since all the inform tion required for practical purposes can often be obtained by studyi the configuration of points on the scatter diagram; for this purpose t co-ordinate system supplied by the centroid factors is as good as a other, since the axes can be forgotten once the points have been plotte Moreover, sometimes the points on the scatter diagram do not fall in

ch obvious groups. Yet when they do we may sometimes wish to ake our factor loadings more descriptive by changing to a different ir of factors which can be interpreted, and yet, being linear functions the centroid factors, convey essentially the same information when nsidered jointly. Thus, for Burt's example, if we write

$$\left.\begin{aligned} i &= 0{\cdot}712l + 0{\cdot}494h \\ u &= 0{\cdot}796l - 0{\cdot}948h \end{aligned}\right\} \quad \ldots\ldots(10)$$

obtain a new specification of the test variable x_1 as follows:

$$\left.\begin{aligned} x_1 &= 0{\cdot}87i + 0{\cdot}41u + \eta_1 \\ &= 0{\cdot}78l + 0{\cdot}23h + \eta_1 \end{aligned}\right\} \quad \ldots\ldots(11)$$

milarly,

$$\left.\begin{aligned} x_2 &= 0{\cdot}72l - 0{\cdot}16h + \eta_2 \\ x_3 &= 0{\cdot}61l - 0{\cdot}07h + \eta_3 \\ x_4 &= -0{\cdot}01l + 0{\cdot}59h + \eta_4 \\ x_5 &= 0{\cdot}01l + 0{\cdot}68h + \eta_5 \\ x_6 &= 0{\cdot}28l + 0{\cdot}41h + \eta_6 \end{aligned}\right\} \quad \ldots\ldots(12)$$

If we plot the coefficients of h against those of l, as in Fig. 2, we see at one axis now passes through the centroid of the literary tests 1, 2, d 3, while the other passes through the centroid of the handicraft

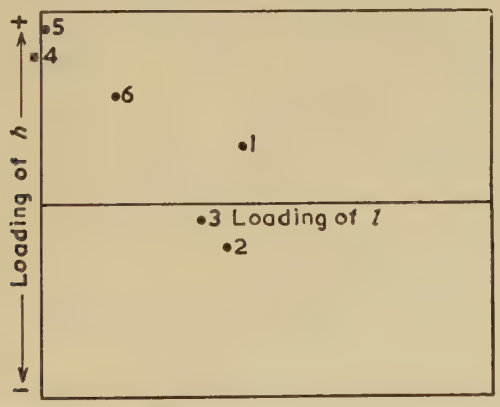


FIG. 2. Burt's example: scatter diagram for factor loadings in the factors l and h.

ts 4 and 5; the former have great loadings in l, which may thus be ntified as a *literary* factor and the latter in h, a *handicraft* factor. ike i and u, which were uncorrelated in the population of candi- es, l and h, though still having unit variances, have a correlation of 5; this is not a serious disadvantage, though it alters the interpreta- n of the angles between the lines joining the test points to the origin, l also distorts the shape of the configuration of scattered points. ese troubles can be avoided by using oblique axes of co-ordinates, lined at an angle $\cos^{-1} 0{\cdot}35$, for the plotting; then the configuration points in Fig. 1 is reproduced exactly, and the axes come in the

position of the broken lines in Fig. 1. The mathematics underlying th choice of the numerical coefficients in (10) is given in the Appendi along with the formulae which are required.

Clearly this final stage in the analysis is not completely objective an unique from the mathematical point of view, depending as it does o the analyst's subjective appreciation of groups of directions on th scatter diagram, which is affected not only by the actual configuratio of points but also by the nature of the tests; different workers wit different attitudes could easily give different but equivalent facto interpretations of the same data. This lack of uniqueness is yet anoth illustration of the principle, familiar in regression analysis, that inte pretation in terms of cause and effect is not strictly a statistical proble

APPENDIX

Changing to an Alternative Pair of Factors

Suppose the equation of the line through the centroid of one grou of tests (e.g. 1, 2, 3) is $\beta = m_1\alpha$, and that through the centroid of th other is $\beta = m_2\alpha$. We wish to replace factors i and u by alternati factors l and h, linearly related to them; let us then put

$$i = \theta' l + \phi' h \qquad u = \theta l + \phi h$$

where θ, θ', ϕ, ϕ', are constants to be determined.

For any test x with loadings α, β,

$$x = \alpha(\theta' l + \phi' h) + \beta(\theta l + \phi h) \qquad \ldots.(1$$

If the test point lies on the first line, the coefficient of h is to be zer hence

$$\alpha\phi' + m_1\alpha\phi = 0, \text{ i.e. } \phi' = -m_1\phi \qquad \ldots.(1$$

Similarly, if the test point lies on the second line, the coefficient of l to be zero; hence

$$\theta' = -m_2\theta \qquad \ldots.(1$$

Since $$\text{var } u = \theta^2\text{var } l + 2\theta\phi\text{cov}(l,h) + \phi^2\text{var } h \qquad \ldots. (1$$

it follows from the definition of all factors as having unit variance th

$$\theta^2 + 2\theta\phi\rho_{lh} + \phi^2 = 1 \qquad \ldots.(1$$

where ρ_{lh} is the correlation coefficient between l and h.

Similarly by considering var i and using (14) and (15),

$$m_2{}^2\theta^2 + 2m_2m_1\theta\phi\rho_{lh} + m_1{}^2\phi^2 = 1 \qquad \ldots.(1$$

and since $\text{cov}(i,u) = 0$ it follows that

$$m_2\theta^2 + (m_1 + m_2)\theta\phi\rho_{lh} + m_1\phi^2 = 0 \qquad \ldots.(1$$

lving (17), (18), and (19) simultaneously for θ, ϕ and ρ_{lh}, we obtain e formulae

$$\theta = \pm\sqrt{(1 + m_1^2)}/(m_1 - m_2) \quad \ldots.(20)$$

$$\phi = \pm\sqrt{(1 + m_2^2)}/(m_1 - m_2) \quad \ldots.(21)$$

$$\rho_{lh} = \pm(1 + m_1 m_2)/\sqrt{[(1 + m_1^2)(1 + m_2^2)]} \quad \ldots.(22)$$

ne signs of θ, ϕ are chosen to give the 'purer' tests positive loadings their dominant factor, and the sign of ρ_{lh} is opposite to that of $\theta\phi$. r Burt's example,

$$m_1 = \frac{\frac{1}{3}(0{\cdot}21 + 0{\cdot}44 + 0{\cdot}33)}{\frac{1}{3}(0{\cdot}87 + 0{\cdot}52 + 0{\cdot}49)} = 0{\cdot}521 \qquad m_2 = \frac{\frac{1}{2}(-0{\cdot}40 - 0{\cdot}45)}{\frac{1}{2}(0{\cdot}43 + 0{\cdot}52)} = -0{\cdot}895$$

$= 0{\cdot}796$, $\theta' = 0{\cdot}712$; $\phi = -0{\cdot}948$, $\phi' = 0{\cdot}494$; $\rho_{lh} = 0{\cdot}35$.

Any other two lines could have been used, of course, and would have l to an alternative (but in this case less practically descriptive) pair factors.

REFERENCES

BURT, Sir CYRIL (1940). *The Factors of the Mind.* University of London Press.

HARPER, R. (1956). 'Factor analysis as a technique for examining complex data on foodstuffs', *Appl. Statist.*, **5**, 32.

THURSTONE, L. L. (1935). *The Vectors of the Mind.* University of Chicago Press.

NOTES AND COMMENTS

Readers will be interested to note that the International Statistical Institu are holding a conference in Brussels from 2nd to 8th September 1958. Th conference is a special session of the Institute, and the topics to be discusse are:

Recent developments in experimental design

Statistical methods for the assessment of living conditions

The use of sampling methods

International aspects of the censuses of population

The use of statistical methods in earth sciences and astronomy

Historical development of statistics

Statistical problems in interdependent dynamic systems

There will also be joint meetings with the Biometric Society and with th European Organisation for Quality Control.

★ ★ ★

We are pleased to hear from time to time of the activities of the Britis Computer Society, which are being widened by the formation of ne branches, the most recent of which we have heard being at Cardiff ar Sheffield. Since the application of statistical methods depends so largely o efficient computation, these activities should be of interest to many of o readers.

★ ★ ★

On the other side of the Atlantic, the Southern Regional Gradua Summer Session in Statistics for 1958 is being held at the Oklahoma Sta University, Stillwater, from 16th June to 25th July. The sponsoring org nisations are Oklahoma State University, University of Florida, Nor Carolina State College, Virginia Polytechnic Institute, and the Southe Regional Educational Board. An impressive panel of lecturers has be engaged and a wide field will be covered.

★ ★ ★

The College of Technology, Birmingham, is again offering a full-ti course of three weeks' duration on 'Statistical Quality Control and Acce tance Sampling', commencing on Monday, 8th September. It will be direct by Dr C. J. Anson, Senior Lecturer in the Mathematics Department, a Mr J. D. Morrison, Senior Lecturer in the Production Engineeri Department. Further details and application forms can be obtained fr the Registrar, College of Technology, Gosta Green, Birmingham, 4.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT GROUP

h December 1957. Estimation Problems in Stochastic Processes: S. K. Zaremba

h January 1958. Non-linear Estimation: G. A. Coutie

h February 1958. Ranks and Variate Values: M. G. Kendall

h March 1958. Random Walk Problems: L. S. Goddard

EDINBURGH GROUP

th January 1958. Operational Research (film strip): A. Stafford Beer. Joint meeting with the Industrial Panel of the Scottish Film Council.

th February 1958. Some Methods for Reducing Uncertainty Concerning the Future: J. V. Gregg

th March 1958. Market Research: J. F. Hayes. Joint meeting with the Scottish Economics Society.

GLASGOW GROUP

th March 1958. Statistical Methods Applied to Attitude Studies: N. J. Squirrell

LEICESTER GROUP

th January 1958. Smoking and Cancer of the Lung: W. R. S. Doll

th February 1958. Machine Interference: F. Benson

h March 1958. Methods of Productivity Measurements: T. E. Easterfield

LONDON GROUP

h December 1957. Half-day Conference on Production Control.

he individual subjects and speakers were: Introductory Remarks on ernetics (A. Stafford Beer); Production Control in a Chemical Factory Battersby); Computing Methods for Optimising Production Plans (C. M. ners-Lee).

February 1958. Statistics in Management and Administration: L. H. C. ippett. Joint meeting with the Royal Institute of Public Administration.

March 1958. A Review of some Sampling Inspection Schemes: I. D. ill

MERSEYSIDE GROUP

11th February 1958. Operations Research in the UK and the US: B. H. Rivett

11th March 1958. Management Accounting: W. J. Jennett

NORTH-EASTERN GROUP

11th December 1957. The Statistical Approach to Stocktaking: E. S. Miln

Mr Milner first explained how the decision to review the organisatio methods, and accuracy of stocktaking arose out of the occurrence of u explained differences in stocks of a certain company. The methods use involved the estimation of the quantities of stock in tanks of greatly varyi sizes, and it was suggested that they should be modified so as to be appr priate to the size of tank. Mr Milner showed how it had been found th the most efficient way to achieve a given accuracy in the estimation of t total stock was to expend most of the total effort on obtaining accura estimates for the largest tanks, to be content with somewhat less accura ones for the medium tanks, and to be satisfied with approximate estimat for the small tanks.

15th January 1958. The Effects of Television: W. A. Belson. Joint meeti with the British Psychological Society.

The basic difficulty in studying the effects of television is that viewers a ordinarily subject to a large number of other influences as well. Dr Belso described the *stable correlate method*, which is a development from the pr cedure in which differences in the behaviour of matched groups of viewe and non-viewers are interpreted as estimates of the effect of television. this method the matching criteria are established empirically; there is discarding of unmatched subjects and there is a built-in check on the co pleteness of the matching achieved. Dr Belson discussed some results bas on the method and described some current research. (A paper dealing wi the stable correlate method was published in *Applied Statistics*, **5**, 195.)

12th February 1958. Colliery Accidents and Statistical Design: F. D. Liddell

Owing to the considerable number of fatal accidents and very lar number of injuries that occur in collieries in Great Britain, statistic research, sponsored by the NCB Medical Service, has been carried out f several years to help to improve the standard of safety. Data analys include: (i) records of all injuries reported in a year at eight collieries; (records of all compensable injuries in all collieries, 1946–49; (iii) records deaths and injuries reportable to the Mines Inspectorate in nearly 2 collieries in 1955. Mr Liddell described this work, some of the difficult encountered and results obtained; he also discussed future projects.

SHEFFIELD GROUP

30th January 1958. The Installation of a Quality Control System: D. Webb.

27th February 1958. Unfair Queues: F. Downton

27th March 1958. Automatic Sales Forecasting: A. Muir

SOUTH WALES GROUP

th January 1958. Simple Experimental Designs: D. R. Cox

th February 1958. The Design of Screening Tests in the Pharmaceutical Industry: O. L. Davies

th March 1958. The Use of a Computer in a Monte Carlo Synthesis: R. Neate

STUDY SECTION

LONDON GROUP

1 January 1958. The Measurement of Non-industrial Work: H. W. G. Deeks

h February 1958. Discriminant Analysis and Perinatal Mortality: J. A. Heady

th March 1958. Planning and Organising Budget Surveys: W. F. F. Kemsley

BRISTOL GROUP

h December 1957. The Full-employment Level: R. J. Pounce

After discussing some of the problems of definition, Mr Pounce explained r methods of analysing local rates of unemployment and vacancies in er to assess the rate of unemployment at which the demand for labour proximates to the supply.

h January 1958. Motivation Research: D. H. Pickard

Mr Pickard defined motivation research as investigation of the sub-scious motives of the consumer and pointed out that the consumer's aviour may be influenced by factors which he either does not know ut or does not wish to reveal. He discussed the interviewing techniques d and outlined the procedure for a motivation research survey, men-ing some of the statistical problems encountered.

h February 1958. Election Forecasting: D. E. Butler

Mr Butler gave a definition of psephology and dealt with laws relating number of votes to the distribution of seats. He explained some of the iculties encountered in using public opinion polls to predict election lts, and compared the results of by-elections with those of general elec-s. Finally he discussed the effects of election campaigns.

h March 1958. The Planning of Experiments: D. R. Cox

Dr Cox enumerated the points to be considered before designing an eriment, and the assumptions made. He then discussed the choice of tments, the observations to be taken, and methods for increasing pre-on. The talk was illustrated by examples from the textile industry, dicine, and psychology.

MEDICAL SECTION

1 January 1958. Mortality of Miners: D. D. Reid

1 February 1958. Pulheems: S. Rosenbaum

1 March 1958. Methodological Problems in the Study of Occupational Mortality: M. A. Heasman

BOOK REVIEWS

Statistical Sources for Market Research. London: The Market Research Society in associati with the Oakwood Press, 1957. Pp. viii + 32. 10s. 6d.

The contents of this excellent source book are difficult to reconcile wi its stated purpose. The standards of membership imposed by the Mark Research Society are strict, and members are of a quality likely to have t personal contacts which for most practising statisticians replace the sour book. Nearly all market research men are members of the Society.

The contents list is based on the economic sections of the *Annual Abstr of Statistics*, comprising population, raw materials and foodstuffs, industr housing, transport, national income and expenditure, and internation trade. The last chapter deals with advertising and readership.

The interpretation of 'market research' is therefore 'desk research' a not field work. A distinguished foreign market research man, using t source book alone, would fare well at some stages and ill at others. If t industry of his interest had a strong Trade Association, had been strict controlled at some time, or had the good fortune to be examined by t Royal Statistical Society, the historical background could be filled in. investigation in the country by sampling techniques were necessary, could formulate a design based on persons, or even families; he cou stratify by employment, co-op. membership, by age at marriage, or newspapers read. He would get no help in choosing his sampling uni since neither the Local Authority rating lists nor the Register of Electo are mentioned. Area sampling would be barred, since no list of maps town plans is given. Yet these are important sampling frames. No guidan would be available on the prices of goods, the Index of Retail Prices bei hidden in the 'Labour' section. He could not find any references to dema elasticities for products consumed in the UK, though these have be covered for foodstuffs by the FAO and others.

Within its limits the coverage is good; the treatment is concise, yet manages to convey the type of warning that results from much study of t small print at the foot of statistical tables. It will clearly become a wide used reference book, and it is to be hoped that its misleading title will n prevent its widespread purchase.

The hard-wearing cover is dark blue in colour, a fact which m explain why the relevant publications of the University of Cambridge a not mentioned.

C. NUTTA

Philip J. McCarthy: *Introduction to Statistical Reasoning.* New York, Toronto and Lond McGraw-Hill, 1957. Pp. xiii + 402. 43s. 6d.

In the recent past, a procession of textbooks giving an introduction statistical methods has emerged from the presses. The authors of such boo must by now be finding it difficult to construct original titles. This m account for the interesting title of this book, for such an introductory te book it is. However, it is perhaps better to give Professor McCarthy t

nefit of the doubt, as it does describe his advertised intent. This is the dable one of laying stress on the basic approach and thinking in modern tistical methods.

This viewpoint gives rise to a book which has a distinct slant, both in atment and content. Certainly the advantages of the bias are debatable, Professor McCarthy goes far towards disarming criticism along these es by giving, in the introduction, a concise statement of the considerations ich guided him in the planning of the book, and an explanation of the sequences. So any discussion must centre around these points, the most portant being:

....(2) Most people who study statistics are not equipped to handle a thematical approach. (3) The use of a non-mathematical approach does absolve one from exercising great care in stating the reasoning and hniques involved in statistics. .....(5) The concepts of statistical soning are novel and difficult for most students in a final course and refore should be introduced gradually and repeated time after time. ...(7) Illustrative material should be drawn from significant investiga-ns, and should not merely be sets of data upon which to perform computa-ns.' Later it is admitted that several important topics are conspicuous their absence, notably small-sample procedures based on the t, χ^2, and listributions, and the analysis of variance.

Other consequences are also apparent; for example, the style adopted is remely discursive. This is not unusual in American textbooks, but here author clearly regards it as a virtue. There may well be others who agree. Is it necessary for clarity and conciseness to be rejected when a n-mathematical' approach is taken? This book certainly deals with its ited compass of subjects in a very thorough and extended way: for mple, thirteen pages are occupied with the definition and calculation he arithmetic mean.

Perhaps this is a trap especially reserved for the lecturer who converts lecture notes into a book. Verbal exposition necessarily involves repeti-n of the argument, since, once spoken, it cannot be recalled in its entirety. wever, the written word can always be re-read until comprehension vns. This means that the author can use a concise statement provided t he also has the skill to make it a clear and efficient communication.

Professor McCarthy teaches at the New York State School of Industrial Labor Relations, Cornell University, and the illustrative examples in book 'range broadly over the social sciences'. Each example is dealt n at some length in an attempt to sketch the background of the work to ch the statistical method contributed. This seems a good idea, but in ctice the excursions are long enough to prove distracting from the main ne, but not really extensive enough to provide an adequate case history. Instance, after the variance and standard deviation have been announced subject, the data to be used as an example are presented. A half-page en devoted to discussion of these data, and it is apparently important is to realise at this stage that 'great care must be used in defining the upations of machine-tool operator and tool-and-die sinker'.

This book, then, does not add competitively to the range of introductory books available in this country, particularly as it covers only a part of really adequate introductory course on statistics.

PHILIP J. STANLEY

Philip Lyle: *Regression Analysis of Production Costs and Factory Operations* (Third Editi revised by Leonard H. C. Tippett). Edinburgh and London: Oliver and Boyd Ltd., 19 Pp. xvi + 204. 16s.

This book, first published in 1944, had its second edition in 1946 and h been out of print for some time. It was known that Philip Lyle intend making revisions for a new edition, but he died before he could underta this work. Mr Tippett has, in preparing the third edition, corrected err in the second edition and inserted a few interpolations in the text. I preface is a useful introduction to the work.

The essence of Mr Lyle's method is to relate variations in cost with var tions in output, using regression technique. Such an analysis leads to estimation of marginal and fixed costs using the parameters of the fitt relationship. He makes a distinction between short-term and long-te cost functions, the latter dealing with mean costs only and ignoring lo fluctuations. The presentation of the method is intimately associated w an exposition of orthodox regression analysis. It may be, as Mr Tipp says in his preface, that the latter are presented in fuller detail than necessa for the purpose in hand, but there is a quality in the author's style a treatment that avoids tediousness. There are two appendices giving inform tion on computational methods, significance testing, curve fitting, furth work on correlation and regression, and nomograms for testing the sign cance of t and r (Mr Lyle was a notable exponent of the use of nomogra in statistical work).

The book, written 'with the main object of trying to interest industr workers in a very useful subject' (original preface), had in mind the accou tant and technologist as readers. It has probably been useful as a prim in regression analysis, but it is doubtful if the cost application has ma much impression on the accountant. This is regrettable and is probal indicative of poor communication between statisticians and accountar It seems unlikely that the method given in the book could be applied wi out modification outside a sugar refinery, as the author would have be the first to admit; but the time may have come, with the publication of third edition, to think again about the possible application of statisti methods to factory costs. If the accountant might benefit from a knowled of statistics, it is equally true that the industrial statistician might ben from a critical appreciation of the meaning of costs—particularly if he engaged in linear programming or other operational research exercises.

W. N. JES

PUBLICATIONS RECEIVED

E. Devons: *An Introduction to British Economic Statistics*. Cambridge University Press, 19 Pp. vii + 255. 22s. 6d. (Reviewed in the article commencing on p. 76.)

Scott W. Hazen, Jr.: *A Comparative Study of Statistical Analysis and other Methods of compu Ore Reserves, utilizing Analytical Data from Maggie Canyon Manganese Deposit, Artil Mountains Region, Mohave County, Ariz.* Bureau of Mines Report of Investigations 5 United States Government Printing Office, 1958. Pp. viii + 188.

Effortless figuring

With a Monroe 8N-213 you get all the advantages of a fully automatic adding-calculator. Drudgery disappears from routine figure work and even the most complex calculations become merely a matter of feeding figures and pressing easy-to-select keys. Exclusive Monro-Matic features mean that there is less to remember, less to do, less possibility of human error. Here is a machine that cuts costs and saves time in every department where figure work counts.

Monroe Calculators range from simple hand-operated models to the world's most advanced machines. With Monroe you can bank on the best advice on choosing and using your machine and the most efficient and reliable after-sales service.

ONROE

For further particulars, please write or telephone:

MONROE CALCULATING MACHINE COMPANY LIMITED.
BUSH HOUSE, LONDON W.C.2. COVent Garden 0211

Titles from an important statistical list

SAMPLING METHODS FOR CENSUSES AND SURVEYS FRANK YATES, F.R.S. 2nd Edn. **42s.**

STATISTICAL METHOD IN BIOLOGICAL ASSAY D. J. FINNEY, F.R.S. **78s.**

THE DESIGN AND ANALYSIS OF EXPERIMENT M. H. QUENOUILLE **42s.**

RANK CORRELATION METHODS M. G. KENDALL 2nd Edn. **40s.**

EXERCISES IN THEORETICAL STATISTICS M. G. KENDALL **24s.**

PROBABILITY AND THE WEIGHING OF EVIDENCE I. J. GOOD **21s.**

BIOMATHEMATICS C. A. B. SMITH 3rd Edn. **80s.**

"GRIFFIN'S STATISTICAL MONOGRAPHS . . . . ."

No. 1 **ANALYSIS OF MULTIPLE TIME-SERIES** M. H. QUENOUILLE **24s.**

No. 2 **A COURSE IN MULTIVARIATE ANALYSIS** M. G. KENDALL **22s.**

CHARLES GRIFFIN & COMPANY LTD

42 DRURY LANE *Established 1820* **LONDON WC2**

FEDERATION OF NIGERIA

Statistician to Supervise Staff Training

Qualifications: Good Honours Degree or other appropriate qualifications in Statistics and appropriate teaching experience.

Duties: The training of Assistant Statistical Officers to final standard of Association of Incorporated Statisticians; the officer may also be required to undertake the ordinary duties of a Statistician in the Department.

Terms: Contract/gratuity. Gross emoluments ranging from £870–£1,824. Free passages; free medical attention. Generous home leave; low income tax.

Application forms and further particulars from Director of Recruitment, Colonial Office, Sanctuary Buildings, Great Smith Street, London, S.W.1. Quote BCD 59/14/02.

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part
20s.

Annual Subscription
Post Free £4 4s.

Principal Contents of Volume 121 Part 1 1958

Recent Developments in Economic Statistics: The Inaugural Address of the President, Sir Harry Campion, C.B., C.B.E., delivered November 20th, 1957 (with Proceedings)

The Depreciation and Mortality of Motor-Cars (with Discussion) .. .. *J. S. Cramer*

The British Propensity to Save .. .. *L. R. Klein*

A Note on the Comparison of Exchange Rates and Purchasing Power between Countries .. .. .. *R. C. Geary*

Generalization of Yastremsky's Formula for Analytical Graduation of Fertility Rates .. .. *Ivo Lah*

Obituary: Roy Goncalves Glenday

Reviews of Books, Statistical and Current Notes, Additions to Library, Periodical Returns

---

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1

Telephone: Welbeck 7638

Journal of the Royal Statistical Society

SERIES B—METHODOLOGICAL

e Each Part 30s.

Annual Subscription Post Free 61s.

Contents of Volume 19 No. 2 1957

BLEMS IN THE PROBABILITY THEORY OF STORAGE SYSTEMS *J. Gani*

E PROBLEMS IN THE THEORY OF DAMS *D. G. Kendall* (With Discussion)

TINE ANALYSIS OF REPLICATED EXPERIMENTS ON AN ELECTRONIC COMPUTER *F. Yates, M. J. R. Healy and S. Lipton* (With Discussion)

IMAX PROCEDURE FOR CHOOSING BETWEEN TWO POPULATIONS USING SEQUENTIAL SAMPLING *R. J. Maurice*

ICIENT STATISTICS, SIMILAR REGIONS AND DISTRIBUTION-FREE TESTS *G. S. Watson*

THE USE OF THE NORMAL APPROXIMATION IN THE TREATMENT OF STOCHASTIC PROCESSES *P. Whittle*

VARIANCE OF THE MEAN OF A STATIONARY PROCESS *E. J. Hannan*

THODS OF CONSTRUCTION AND ANALYSIS OF SERIALLY BALANCED SEQUENCES *M. R. Sampford*

MODIFIED LATIN SQUARE *B. Rojas and R. F. White*

LYSIS OF VARIANCE AS AN ALTERNATIVE TO FACTOR ANALYSIS *M. A. Creasy*

E FURTHER RESULTS IN THE NON-EQUILIBRIUM THEORY OF A SIMPLE QUEUE *N. T. J. Bailey*

MAL APPROXIMATION TO MACHINE INTERFERENCE WITH MANY REPAIR MEN *P. Naor*

TIONARY DISTRIBUTIONS OF THE NEGATIVE EXPERIMENTAL TYPE FOR THE INFINITE DAM *J. Gani and N. U. Prabhu*

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1

Telephone: Welbeck 7638

BIOMETRIKA

VOL. 45, PARTS 1 AND 2 CONTENTS JUNE 19

ARMITAGE, P. Numerical studies in the sequential estimation of a binomial parameter-
LESLIE, P. H. A stochastic model for studying the properties of certain biological systen
by numerical methods—SKELLAM, J. G. On the derivation and applicability of Neyman
Type A distribution—BLISS, C. I. and OWEN, A. R. G. Negative binomial distributio
with a common k—BRASS, W. Simplified methods of fitting the truncated negati
binomial distribution—COX, D. R. The interpretation of the effects of non-additivity
the latin square—ASHFORD, J. R. Quantal responses to mixtures of poisons under cond
tions of simple similar action—the analysis of uncontrolled data—MOORE, P. G. Son
properties of runs in quality control procedures—WILLIAMS, E. J. Simultaneous regre
sion equations in experimentation—RUSSELL, T. S. and BRADLEY, R. A. One-w
variances in a two-way classification—PLACKETT, R. L. Studies in the history of prob
bility and statistics. VII. The principle of the arithmetic mean—BROWN, R. L. ar
FEREDAY, D. Multivariate linear structural relations—OLKIN, I. Multivariate rat
estimation for finite populations—BARTON, D. E., DAVID, F. N., and MALLOWS, C.
Non-randomness in a sequence of two alternatives: I. Wilcoxon's and allied test statisti
—GOODMAN, L. A. Simplified runs tests and likelihood ratio tests for Markoff chains-
LEIPNIK, R. B. Moment generating functions of quadratic forms in serially correlat
normal variables—SAW, J. G. Moments of sample moments of censored samples fro
a normal population—HERDAN, G. The relation between the dictionary distribution ar
the occurrence distribution of word length and its importance for the study of quantit
tive linguistics—MACKENZIE, J. K. Second paper on statistics associated with t
random disorientation of cubes—WAUGH, W. A. O'N. Conditioned Markov processes-
MISCELLANEA: Contributions by D. E. BARTON, D. E. BARTON and F. N. DAVID,
BERGSTROM, C. W. CLUNIES-ROSS, H. R. B. HACK, M. HALPERIN and S. W. GREENHOUS
G. HERDAN, A. HUITSON, N. L. JOHNSON, J. E. KERRICH, R. D. LORD, R. MAURICE-
CORRIGENDA: F. N. DAVID and M. G. KENDALL, A. J. MAYNE, E. L. CROW—REVIEWS-
OTHER BOOKS RECEIVED.

The subscription, payable in advance, *is* 54*s.* (*or* $8) *per volume* (*including postage*). *Cheques should made payable to* Biometrika, *crossed* "a/c Biometrika Trust" *and sent to the Secretary, Biometrika Offi Department of Statistics, University College, London, WC*1. *All foreign cheques must be drawn on a ba having a London agency.*

Issued by THE BIOMETRIKA OFFICE, University College, London

Scottish Journal of Political Economy

The Journal of the Scottish Economic Society

Contents of Volume V, No. 2
June 1958

Symposium on Wages Policy

Wages Policy and Inflation .. .. *A. K. Cairnc*
Wage Stability and Employment .. .. *J. R. Parkin*
The Inadequacy of Recent Wage Policies in Britain .. .. *D. J. Rober*
Can Britain have a Wage Policy? .. .. *A. Flan*
Wage Policies Abroad .. .. *T. L. Johns*

Comments:

The Conditions for the Avoidance of the Spiral .. .. *E. H. Phelps Br*
Two Schools of Thought on Wage Inflation .. .. .. *H. G. Johr*
Centralized Wages Policy: A Comment .. .. *B. C. Rob*
Wages and Demand .. .. *A. J. Br*
Wages Policy: A Comment .. .. *D. T. J*
Wages and Prices .. .. *A. D. Camp*
The Inadequacy of Recent Wage Theorising .. .. .. *H. A. Tu*
Organising an Opposition to Price Inflation .. .. .. *J. C. R. I*

The SCOTTISH JOURNAL OF POLITICAL ECONOMY is published three times per year, an free to members of the Scottish Economic Society. The annual subscription to non-memb is £1 10s. (U.S.A. and Canada $5). Single copies 10s. 6d. (U.S.A. and Canada $1. post free. Orders should be sent to Oliver and Boyd Ltd., Tweeddale Court, High Str Edinburgh, 1.

A GUIDE TO ACCEPTANCE SAMPLING

LEONARD H. C. TIPPETT
Shirley Institute, Manchester

This is an expository article, more for people who have to use acceptance sampling than for those who have to develop schemes in detail. Readers who are not familiar with the subject should find Mr Tippett's article a useful introduction to Dr Hamaker's, which appears elsewhere in this issue.

CEPTANCE sampling is the procedure adopted for taking samples
n batches or lots of some industrial product, testing or inspecting
samples, and on the basis of the results deciding whether the whole
ches are satisfactory. The literature on the subject is extensive,
ttered, and somewhat confusing. Some papers and books develop
ie new procedure or set of tables and are primarily concerned with
ical and mathematical details; these are difficult reading for the
xpert. Some present tables and procedures almost as cookery-book
pes; these are very practical but not satisfying to anyone who is
villing to accept everything on trust. Some text-books have excellent
ositions but none that I know of covers the ground completely. In
ious places, notably in *Industrial Quality Control*, the journal of the
erican Society for Quality Control, there are articles describing the
lication of acceptance sampling to particular industrial situations.
his expository article attempts to survey at least the main well-
blished methods of acceptance sampling by dealing with broad
nciples. It is primarily intended for the industrialist who wishes to
n a general appreciation of the subject and the student who wants
ackground against which subsequent studies and reading can be seen.
aders will not be assumed to have much acquaintance with statistics
ond the general ideas of probability, averages, and variability, and
principles of their measurement.

Acceptance procedures are adopted when an industrial product
ses from the responsibility of those who supply it to that of those
o are to use it—from the 'producers', as they are technically termed
he literature, to the 'consumers'. The product under discussion
be a raw material, a manufactured or partly manufactured com-
ent, or a finished thing. The producer may be a merchant, a
ufacturer independent of the consumer, or a department in the
e factory as the consumer. The consumer may be a factory or
artment that is going to process the product further or the ultimate
sumer in the ordinary sense of the word.

Most of the literature on acceptance sampling concerns products that
discrete articles, and little has been written about continuous products

such as liquids and powders. Only discrete articles are considered her and what is written applies to liquids only if they can be considered discrete units such as cans-full, to powders if they can be considered sacks-full or packets-full, and so on.

Industrial products may be appraised by some measurement, perha of dimension, or life, or strength. This is termed appraisal by 'variabl and the quality of a collection of the articles is described by such quan ties as the average and the standard deviation or mean range. T other method of appraisal is to class each article as satisfactory or u satisfactory according to some technical criterion—whether or not dimensions come within the tolerance limits of some specified valu whether or not the surface finish is satisfactory, whether or not electrical component passes some insulation test, and so on. This termed appraisal by 'attributes' and the quality of a collection articles is described by the fraction or percentage of satisfactory unsatisfactory articles present. Unsatisfactory articles are almost u versally shortly termed 'defectives' and the fraction or percentage defectives describes the quality of the collection. It is unfortunate, a to the tyro somewhat disconcerting, that as this measure of quali increases, the quality of the collection in the ordinary sense of the wo becomes worse. However, the convention is too well established any change to be made now, and the reader must be alert to understa the meaning intended for the word 'quality'.

In between appraisal by variables and attributes comes appraisal the number of defects. For the articles under examination, defects m be classed into several types so that any one article may have seve defects. Then the quality of a batch of given size may be specified the total number of defects, or the mean number of defects per artic Readers should be alert to distinguish between defectives and defec Defects are dealt with only occasionally in the literature and will n be further mentioned here.

Acceptance sampling procedures have been most highly develop for appraisal by attributes and most of the literature is about th Accordingly this article will deal at first with percentage defectives a variables will be dealt with shortly at the end.

For some of the quantities that arise different terms are used in Gr Britain and the United States. Where this occurs I shall give bo terms at the first mention (their equivalence is obvious) but adopt t British usage except where this is impracticable. As a start readers m note that the aggregates in which the products are presented acceptance are termed 'batches' in Great Britain and 'lots' in the US

Sampling sometimes presents practical problems such as that obtaining representatives of the articles in the interior of a package a more generally, that of securing samples that are free from bi Economies can sometimes be achieved, too, by taking account of sys matic variations in quality in different parts of the bulk that is be sampled. These matters are a big subject in themselves, especially wh

appraisal is by variables, but I shall not deal with them here. I l assume that the problems of securing samples have all been solved shall deal with what happens after the samples have been obtained.

tudes to Sampling Inspection

he attitude to the use of samples of the industrialist who is unin-cted in the subject—we may term it the naïve attitude—is apt to one of two extreme forms. Some industrialists say that they cannot rd to risk making the mistakes that are inevitable if samples are i: that they can be content with nothing less than 100 per cent ection. Doubtless such an attitude is sometimes justified, but such ustrialists can sometimes be persuaded that 100 per cent inspection have its errors and their associated risks of mistaken decisions about batch, that the risks associated with sampling can be calculated and rolled, and that it is often economical and justifiable to run ulated risks and use samples.

he other extreme, uninstructed, attitude is that of the industrialist has based his decisions on samples for years (perhaps the des-tiveness of the tests on the articles compels him to) and is unconscious sampling errors exist. Such an industrialist will often use samples are hopelessly inadequate in size; he must be persuaded that pling errors exist, and that the associated risks should be calculated taken into account.

r B. P. Dudding[1] has written a pamphlet which is a corrective to attitude. He examines typical sampling clauses in specifications rding to which a batch is regarded as conforming, or not conform-to the specification according to the number of defective articles d in samples of ten or twenty articles; and shows how the operation ch clauses can result in fairly high probabilities of accepting batches high fractions of defective articles and rejecting, as not conforming, hes with low fractions of defectives—i.e. of coming to wrong deci-s. Dr Dudding goes on to suggest how the situation may be roved. It is almost common sense that larger samples will reduce ng decisions.

ne step forward has been taken if calculations of the kind Dr ding illustrates are made and in the light of them different sample are considered, even if the resulting choice is a compromise, ily weighted by what a conservative industry will accept and ely half-understood by the industrialists.

matic Presentation of Sampling Risks

e first step in the systematic presentation of sampling risks is to e a sampling 'scheme' or (in the USA) 'plan'. In the most ele-ary form of scheme, the so-called 'single-sampling' schemes, the ber in the sample from each batch is defined, together with a ber termed the 'acceptance number'. The rule is then stated as

'accept the batch if the number of defectives in the sample equals th acceptance number or fewer, otherwise reject it'.

What happens to the batch after it is 'accepted' or 'rejected'—i rejected, for example, whether it is sent back to the producer or use by the consumer with a rebate on the price—does not concern us.

For any given scheme, with its stated sample size and acceptanc number, the probability of a batch with a given percentage of defectiv articles being accepted can be calculated from statistical theory. Whe these calculations are made for batches with a full range of percentage of defectives from zero to one hundred are plotted, and the points ar joined by a smooth curve, the result is a so-called 'operating character istic' (OC) curve as illustrated in Fig. 1.

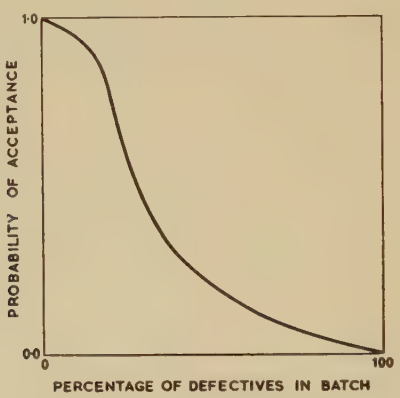


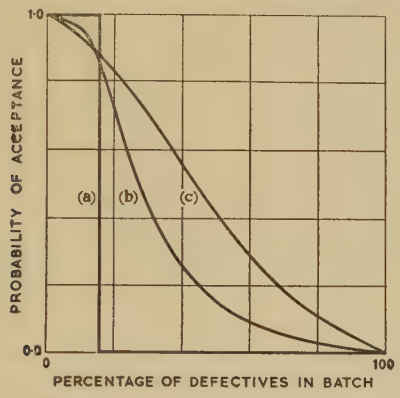
FIG. 1. Typical operating characteristic curve.

If a batch has no defectives, every sample from it will have no de fectives and it will certainly be accepted, so the probability on the O curve is 1·0. If a batch has 100 per cent defectives it will certainly b rejected, and the probability on the OC curve is zero. In between, th probability takes an intermediate value, being high if the percentag of defectives in the batch is low—i.e. if the quality is high.

The probabilities of the OC curve are interpreted in the usual wa as the long-run proportions of batches that will be accepted. If, fo example, the probability of acceptance is 0·86 for a batch percentage defectives of 16, that is interpreted to mean that in the long-run, if th batches with 16 per cent defectives are submitted to the sample test, fraction 0·86 (or 86 in a hundred) of them will be accepted and 0·1 will be rejected.

The object of any sampling scheme is to discriminate between sati factory and unsatisfactory batches, and the OC curve complete describes the power of the scheme to make this discrimination. If w have an album of schemes and corresponding OC curves, and ca choose a curve that describes the required degree of discrimination, w thereby choose the appropriate scheme. If, for the sake of argumen we say that batches with up to 16 per cent of defectives are satisfacto

those with more than 16 per cent are not, 'curve' (a) in Fig. 2 gives ect discrimination at the required level; such would be achieved by the inspection of 'samples' equal in size to the batches. The pling scheme corresponding to curve (b), with its relatively steep gives better discrimination than the scheme corresponding to curve and would, of course, involve taking larger samples. OC curves



2. Operating characteristic curves for schemes having three different degrees of discrimination between satisfactory and unsatisfactory batches.

e to be chosen not only according to their steepness but also accordto the batch percentage of defectives at which the steep part occurs ie percentage at which discrimination is required to be made. arly, OC curves are esoteric things, too complicated to be usable in ctical affairs. Some simplification of description and codification of process of choice is necessary.

lucer's and Consumer's Risks

)ne procedure postulates rather narrow-minded producers and coners whose imagined interests are protected independently. It is in interest of the producer to ensure only that an undue proportion of sfactory batches are not rejected; he does not mind how many unsfactory batches are accepted. Likewise, it is in the interest of the sumer to ensure only that an undue proportion of unsatisfactory hes are not accepted; he does not mind how many satisfactory hes are rejected. *The producer and consumer so described are notional*; producer or consumer in real life can define his interest so narrowly. ccordingly, the notional producer states a percentage of defectives e batch that represents a lower limit of 'goodness' and the risk of tion he is prepared to run for batches at this limit; the percentage fectives is termed the risk point. Then, any sampling scheme that an OC curve passing through the chosen producer's risk (or prolity of rejection) and risk point protects the producer's interests. point is marked in Fig. 3. For example, if the producer chooses a point of six per cent defectives and a rejection risk of 0·05 (probty of acceptance 0·95), he knows that a proportion 0·05 of the

batches with six per cent defectives will be rejected; of batches bette than this, a smaller proportion will be rejected; but of the batches wit more than six per cent defectives more than 0·05 will be rejected. An sampling scheme that gives these results will satisfy him.

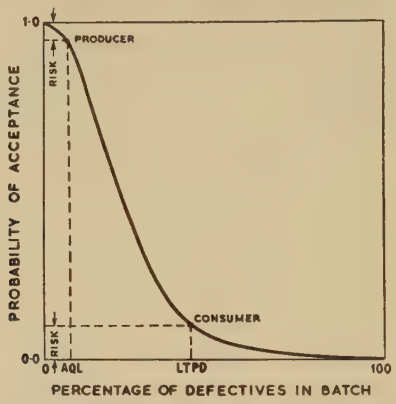


FIG. 3. Operating characteristic curve with consumer's and producer's risks and risk poin indicated, together with the acceptance quality level (AQL) and lot tolerance per ce defectives (LTPD).

The notional consumer states a percentage of defectives in the batc that represents his limit of 'badness' and the risk of acceptance he prepared to run for batches at this limit; the percentage of defective is the consumer's risk point. Any sampling scheme that has an O curve passing through this point (marked in Fig. 3) satisfies him. Thu the consumer's risk point might be 45 per cent defectives and his ris (the probability of acceptance) 0·10. Then a scheme satisfying th criterion will ensure that if there are more than 45 per cent defective in a batch the probability of its acceptance will be less than 0·10; th is the extent of his protection against accepting 'bad' batches.

For a sampling scheme to satisfy both producer and consumer it mu have an OC curve passing through both points; but the precise shap of the curve is of no consequence. There are tables from which th particulars of sampling schemes corresponding to various producer and consumer's risk points and risks can be derived. These will b surveyed in the next section of this article.

It is helpful to notice that two quantities are required to define single-sample scheme—the sample size and the acceptance numbe and the two points on the OC curve provide the two conditions r quired, mathematically, to determine the quantities. It will be see later that other conditions can be specified as alternatives.

The choice of the risk points and risks is, of course, a technical an business choice rather than a statistical one. The quantities are mo tangible to the industrialist than an OC curve and the correspondi choice is doubtless easier to make. But it is not easy, and yet it is n unimportant. Suppose, for example, that the producer's risk point 1 per cent of defectives—perhaps he is fairly confident that he ca

oduce batches that are at least as good as this, and these are the ches that he wants to be fairly sure will be accepted. Further, sup- se that the consumer's risk point is 5 per cent of defectives—batches h up to 5 per cent are tolerable to him but he wants to be fairly sure t batches worse than this will be rejected. One can imagine indus- alists regarding it as being well within their competence to make such ds of choice, even though some difficulty may be experienced. But ficulties of another kind arise in choosing the risks. If they want a h degree of assurance both producer and consumer might choose s as low as 0·01 (the probability of acceptance at the producer's risk nt being 0·99). To give this assurance samples would need to have a e of about 350. If, perhaps for the sake of economising in inspection, producer is content to run a risk of 0·05 and the consumer one of 0 (for some reason, consumer's risks are usually greater than the oducer's), the sample size can be reduced to about 80. It must be y hard for the industrialist to choose rationally between such levels risk. Yet such choices have been made, and convention and experi- e have established a few levels as reasonable. Risks at the lower of above-mentioned levels of assurance (0·05 and 0·10) are commonly d as the basis.

To what extent in practice sampling schemes are based on producer's d consumer's risks and risk points I do not know. Reading of the rature gives one the feeling that the whole conception, fundamental ugh it is to thinking about sampling risks, is rather academic; the mples so often seem to be artificial. However, my experience in this d is slight and I hope that someone with more experience will be e to give readers of this journal better information.

The next section surveys some published sets of schemes based on the ception of this section.

emes derived from Producer's and Consumer's Risks

n addition to the single-sample type of scheme so far mentioned re are double-sample, multiple-sample, and sequential types.

According to the double-sample type, a first sample of a certain size is en; if it contains *a* defectives or fewer the batch is accepted, if *r* ectives or more it is rejected, otherwise a second sample of the same different size is taken; if, when the second sample is taken, there are efectives or fewer in the two samples combined the batch is accepted, erwise it is rejected. Thus, a decision may require one or two ples.

n extension of this idea to the taking of up to three or more samples lts in multiple-sample types of scheme, and the further extension esting for a decision successively as each individual article is taken lts in a sequential type of scheme. A sequential scheme is a multiple- ple scheme with sub-samples of one article. In some writings ltiple and sequential schemes are not distinguished but the distinction asy to maintain and is worth while.

These three types of scheme are progressively more economical tha single-sample schemes in that they require, on the average, smalle samples for a given power to discriminate between satisfactory an unsatisfactory batches. Moreover, the double-sample type recommend itself to some industrialists because it appears to give a doubtful batc a second chance. (In fact, of course, the ultimate probability of re jection can be made the same for any type of scheme). On the othe hand, the more elaborate schemes are more complicated to work an sometimes present administrative difficulties. The differences betwee the types are not the subject of this article, the similarities are.

The basic similarity is that for a scheme of any type an OC curve showing the probability of ultimate acceptance of the batch, can b calculated. And a scheme to give the required power of discriminatio can be chosen in the way described for the single-sample type. For double-sample scheme five quantities require to be defined, and for multiple-sample scheme many more. Clearly, there is scope here fo considerable mathematical ingenuity and elaboration in establishing according to various criteria, optimum values for the quantities. I practice there is no great elaboration. The second sample in a double sample scheme is usually either equal to the first or twice that size, an the successive samples in a multiple scheme are usually equal. Th producer's and consumer's risks and risk points are calculated and eac scheme is regarded as equal in discriminating power to a single-sampl scheme having the same risks and risk points.

Mr Paul Peach[2] gives tables for determining single-sample, double sample, and sequential schemes for producer's and consumer's risks 0·05. Mr G. Horsnell[3] gives tables for determining single-sampl schemes for producer's risks of 0·01 and 0·05, and for consumer's risk of 0·01, 0·05, and 0·10. Mr F. E. Grubbs[4] gives tables for single sample schemes for a producer's risk of 0·05 and a consumer's risk 0·10. Mr F. J. Anscombe[5] gives tables for sequential schemes and th corresponding percentage defectives at the 0·99, 0·90, 0·50, 0·10, an 0·01 probability points on the OC curve. An interesting variation proposed by Mr A. Golub,[6] who gives charts for determining singl sample schemes when the sample size and the producer's and consumer risk points are given, and the sum of the two risks is minimised.

Schemes for Regular Supplies

Sampling schemes so far described lead to a decision on each batc separately and are applicable, in conception if not always in practic where nothing is known about the sort of quality of products th producer supplies. The consumer's interests are protected by the choi of a level of quality (the consumer's risk point) below which the quali of no single batch is likely to fall.

In many industrial situations, however, the consumer receives su plies regularly from a producer. Broadly he is satisfied with the quali of the regular run of batches supplied (if he were not he would go

ther producer) and is well served if the sampling scheme gives the ducer a strong enough incentive to maintain, or improve, his quality. y scheme for which the producer's risk point for a risk of say 0·05 responds approximately to the tolerable percentage of defectives, this property; for the OC curve is steep in this region and deteriora- n of quality beyond this point rapidly leads to an undue proportion batches being rejected. This, if it occurred, would make life so icult for both producer and consumer that either the producer would ore his quality or the consumer would change his supplier. Accord- to this idea, the function of the sampling scheme is not to separate sfactory from unsatisfactory batches but to discipline the producer l distinguish between satisfactory and unsatisfactory producers. In eral, schemes of this kind give adequate control of quality with ller samples than schemes for assessing batches independently, since y make use of the information derived from runs of batches. There two sets of schemes, based, or apparently based, on something like idea.

he first set is that of the US Army Ordnance Department widely wn as Military Standard 105A.[7] According to the system of this set uantity termed the 'acceptable quality level' (AQL) is specified, in official use by the consumer—the US army authorities. It repre- s the quality that the consumer wants, on the average and in the g run. Then, from the tables provided, the size of sample and eptance number are read for a single-sample scheme appropriate to iven batch (or lot) size and the specified AQL, corresponding to t is termed 'normal inspection'.

he scheme is then operated until results are available for ten con- tive batches from the same producer; from these results the 'process rage' percentage defective is calculated for that producer. If the cess average is within a certain range of the AQL, the range being ted to the error with which results for ten batches estimate the 'true' cess average, normal inspection continues. If the process average is ond the upper limit of this range (i.e. if the quality is significantly se than the AQL), 'tightened inspection', involving larger samples, rescribed. If the process average is beyond the lower limit of the ge (i.e. if the quality is significantly better than the AQL), 'reduced ection' is prescribed. The tables give the details of schemes for tened and reduced as well as normal inspection, and there are ise rules for changing from one level to another if the quality of the lies changes. Tables are also given for corresponding double- and riple-sampling schemes.

he official Standard presents concisely only the information neces- for operating the schemes. This is an advantage to the industrialist it gives the system an arbitrary or 'rule-of-thumb' appearance. OC curves of the schemes are given and from these it may be seen the probability of acceptance when the batch is at the AQL is ut 0·8 for the smallest batches and rises above 0·99 for the largest

batches; the AQL thus corresponds roughly to the producer's risk poin of other schemes, with a producer's risk decreasing as the batch siz increases. The consumer's risk point for a risk of 0·10 is many (abou seven) times greater than the AQL for small batches and for larg batches may be two or three times greater; the explicit protection fo the consumer is therefore slight. Mr H. R. Bellinson[8] has given some thing of the background to the development of Military Standard 105A and it appears that experience largely guided the levels set for normal reduced, and tightened inspection, and, we may infer, that statistica analysis guided the expression of that experience in a self-consistent se of schemes.

It will be noted that there is a two-fold incentive to the producer t maintain or improve his quality; as the process average becomes lowe (corresponding to improved quality), not only does the proportion o batches rejected become less, but the amount (and hence cost) o inspection may be reduced—it seems to be tacitly assumed that in spection is a cost to the producer.

The second set of schemes similar in general kind to those just des cribed are those devised by the so-called Statistical Research Grou (SRG).[9] This Group consisted of a number of American statistician who worked at Columbia University between 1942 and 1945 on defenc contracts. The SRG schemes are so arranged that batches having a percentage of defectives at the acceptable quality level (AQL) have a probability of 0·95 of being accepted—i.e. the SRG AQL correspond explicitly to the producer's risk point for a risk of rejection of 0·05 Corresponding to the three levels of inspection (normal, reduced, an tightened) of Military Standard 105A the SRG schemes have five designated I, II, III, IV, and V. These are chosen so that level II involves approximately the same amount of inspection as the norma level in the schemes of Military Standard 105A, and the others involv (in order) one-half, three-quarters, one-and-a-half times, and twice tha amount. The statistical basis of these levels is not stated. Single- double-, and multiple-sampling schemes, together with the corres ponding OC curves, are given for various ranges of AQL and batc size, and for the five inspection levels. There is a discussion of th general considerations that should determine the choice of the level o inspection, but the subject seems, to me at least, to be left rather 'in th air', and no precise rules are given.

The Military Standard 105A and SRG sets of schemes comman respect despite the element of arbitrariness they apparently contai They have been devised by statisticians experienced in the field o acceptance sampling and have held their place for a decade. We ma reasonably assume that they work satisfactorily, at least in industri situations similar to those surrounding the supply of ordnance stores.

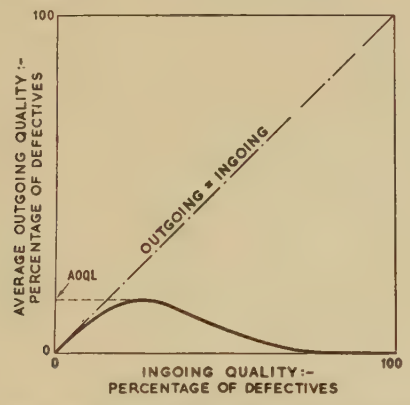
Rectification Schemes

The classic work that finds its ultimate expression in the tables

H. F. Dodge and Dr H. G. Romig[10] introduced a new idea and a
v set of quantities into acceptance sampling. The situation envisaged
he inspection of a long series of batches, particularly such as would
ur if components are passed from one department of a factory to
ther in the course of the manufacture of some assembly, or are passed
ween two closely associated factories. The new idea is that batches
uld not be rejected, in the ordinary sense of the word, but if they
not accepted should be 'rectified' by being inspected completely
all defectives being either removed or corrected.

hus, if the scheme is based on a single sample, each batch has a
ple inspected. If there are fewer than a certain number of defectives
he sample the batch is accepted as it stands, with the defectives it
pens to contain; if there are more than the certain number of
ctives in the sample every item in the batch is inspected, defectives
eliminated, or corrected, and the batch is passed forward perfectly
from defectives.

One new quantity that arises is the average outgoing quality (AOQ).
nsider batches having a given percentage of defectives when pre-
ted for inspection—this percentage is termed the 'ingoing quality'.
a given sampling scheme, a calculable proportion of these will be
epted without change, the remainder will be rectified and will be



4. Typical curves showing, for a rectifying scheme, the relation between the average
tgoing and ingoing quality. The average outgoing quality limit (AOQL) is shown.

sed forward with zero defectives. The average percentage of de-
ives in these batches after inspection is the AOQ; and the AOQ is
ays better (corresponding to fewer defectives) than the ingoing
ity. It is not difficult to see that if, for the given scheme, the AOQ
lculated for a series of values of ingoing quality and the results are
ed, the graph will be something like the curve in Fig. 4. The curve
everywhere be below the line representing the equation:

$$\text{average outgoing quality} = \text{ingoing quality}.$$

low values of ingoing quality few batches are rectified and the
Q is little lower than the ingoing quality. For high values of ingoing

quality a large proportion of the batches are rectified and made perfect and the AOQ becomes very low again. In between, the AOQ rises to a maximum known as the 'average outgoing quality limit' (AOQL). For a given sampling scheme, the average percentage of defectives in outgoing batches cannot exceed the AOQL whatever the ingoing quality.

Each rectifying sampling scheme has its characteristic AOQ curve, just as it has its OC curve (with non-acceptance corresponding to rectification instead of rejection), and the AOQL is one criterion of the degree of control given by the scheme which has proved to meet the requirements of many industrialists.

The AOQL alone, providing only one condition, is not enough to define a single-sample scheme. The other condition underlying the Dodge-Romig tables is that the minimum amount of inspection shall be done—the amount being the inspection of the samples plus the 100 per cent inspection of the rectified batches. This minimum depends on the process average.

If the process average is known and the AOQL can be chosen (the choice is of course the prerogative of the industrialist), the details of the appropriate scheme for batches of a given size can be read from the tables. If the process average is not known, any of the schemes for the chosen AOQL may be used; the AOQL will be maintained but the sampling inspection may not be the most economical that is possible.

The Dodge-Romig tables also use as an alternative criterion of choice of scheme a consumer's risk point derived from the OC curve. The consumer's risk is arbitrarily chosen as 0·1 (Dodge and Romig have obviously found that to be agreeable to the industrialists with whom they have had to deal), and they term the corresponding risk point the 'lot tolerance percent defective' (LTPD, see Fig. 3). Since there is no rejection of batches there is no interest in the producer's risk and risk point, and the other limiting condition for defining each scheme is again the minimisation of the total amount of inspection. Thus, the alternative set of Dodge-Romig tables utilise LTPD instead of AOQL.

These tables also present corresponding double-sample schemes, both with AOQL and with LTPD as a criterion.

Any sampling scheme can be used as a rectification scheme (provided that if the word 'reject' is used or implicit in the rule the word 'rectify' is substituted) and an AOQL can be quoted. The tables of the Military Standard 105A quote values of the AOQL, and the Statistical Research Group schemes are classified alternatively according to the AOQL or the AQL already described.

Mr D. J. Greb[11] has published a convenient chart and table for determining, from the AOQL, single-sample rectifying schemes that minimise the total amount of inspection. Mr F. J. Anscombe, in the paper already quoted,[5] gives tables for multiple rectifying schemes. One condition is fixed by a form of consumer's risk and risk point, the risk being 0·1 or 0·01 that a given number of defectives remain in the batch

ter inspection. The other condition is either that the total amount of spection is a minimum, or that the AOQL has a chosen value.

Iost Economical Sampling Schemes

In discussions on the application of the foregoing schemes, particularly those involving the choice of a producer's or consumer's risk, ention is often made of the general considerations that govern the rious choices to be made—for example, the consumer's risk should very low if the acceptance of unsatisfactory batches could cause uch damage. Attempts have been made to give these considerations uantitative expression by dealing with the various costs involved.

The general approach of these attempts involves knowing the frency distribution of the quality (percentage defectives) of the batches resented by the producer—termed the 'process curve' or 'process haracteristic'. The average of this distribution is the process average reviously encountered. If production were in the strictest possible ntrol, there would be no variation in quality from batch to batch d no process curve; but there would also be no need for sampling spection of the batches. It is presumed, however, that the quality of oduction is so far stabilised that a stable process curve can be identified.

From the process curve, and from the OC curve of a given sampling heme, it is easy to calculate the fraction of defective items that are cepted and the fraction of effective (i.e. satisfactory) items that are jected. Consider, for example, batches with between $4\frac{1}{2}$ and $5\frac{1}{2}$ per nt defectives (centred on 5 per cent) and suppose that according to e process curve these represent one-sixth of the total batches. Suppse, further, that according to the OC curve for the given scheme the obability of acceptance of a batch with 5 per cent defectives is 0·85. hen, of the total batches submitted, a fraction $\frac{1}{6} \times 0{\cdot}85$ will have 5 per nt defectives and will be accepted; and of the total *articles* submitted, ese accepted batches will contain a fraction $\frac{1}{6} \times 0{\cdot}85 \times 0{\cdot}05$ defective. alculations of this sort done for all percentages of defectives covered the process curve, and summed, and multiplied by the batch size, ve the total defective articles per batch that are accepted. Complementary calculations give the total effective articles per batch that are ected.

Next, we need to know the costs per article associated with these ong decisions—the 'decision costs'. These depend on what is done th rejected effective articles—whether, for example, they are scrapped ent back to the producer—and on the damage that can result from utilisation of accepted defective articles. The decision costs per icle are for the accountants and engineers to determine—perhaps by imates involving a large element of guesswork. These costs per icle multiplied by the numbers per batch calculated according to the evious paragraph give the decision cost per batch resulting from the eration of the sampling scheme. Generally, the larger the sample the aller is the decision cost.

Further, we need to know the sampling cost. This may be so-muc as an 'overhead', plus so-much per article for drawing it from the batch plus so-much per article for inspecting it. Then, since the sample siz is defined by the scheme, the total sampling cost per batch can b calculated.

These two, the decision and sampling costs, are, in principle, cal culated for a range of schemes involving samples of different size, an the scheme for which the sum of the two costs is a minimum is the mos economical scheme.

This, of course, is only a broad statement of the main general line o approach; in the various papers on the subject there are elaboration in detail and slight variations in principle. Dr H. C. Hamaker[12] ha given a good summary exposition of the subject, and Mr F. J. Ans combe[13] an exposition with an example. Mr J. Sittig[14] and Mr W Weibull[15] have developed procedures and equations in some detail an give some tables and charts. Professor D. G. Champernowne[16] ha given equations and tables for determining most economical sequentia schemes. Mr G. Horsnell[17] has adopted a somewhat different approac from that described so far and has given tables for the application of hi methods. He deals with situations in which the process curve denote substantially better quality than that at the consumer's risk point, an assesses the decision cost as being to all intents and purposes the cost o manufacture and inspection of articles that are rejected and no available for use.

A more detailed survey of this subject is given by Dr H. C. Hamake in another article in this issue of *Applied Statistics*.

It seems to me that the attempts that are being made to put accep ance sampling on an economic basis are important. Control of qualit by inspection is important and inspection often adds substantially t manufacturing cost; operating efficiency requires that the economicall optimum degree of sampling should be adopted. Moreover, the variou decision costs, difficult though some of them may be to determine, ar closer to the kind of data that industrialists can supply on a rationa basis than are the various risks that have been described. It would see that schemes based on rough estimates of costs, or even on informe guesses, are likely to be more satisfactory than those chosen in oth ways.

Appraisal by Variables

Where the quality of the individual article can be measured, tha method of appraisal has the advantage that for a sample of given siz there is much better discrimination between satisfactory and unsati factory batches than when the individual articles are merely classifie as defective or effective. The measurement contains more informatio Sometimes the gauging or other process for separating defectives fro effectives is so much quicker and more economical than the process measurement that the method of attributes is more economical in spi

equiring larger samples for a given degree of discrimination. ertheless, measurements are much used in industry.

n view of this it is perhaps surprising that much less has been written acceptance sampling by variables than on that by attributes, and generalised schemes comparable to those described in this article e not been worked out. Whether a sample of measured articles is racterised by a mean, a standard deviation, or some other statistical sure, a single, double, multiple, or sequential scheme can be stated, its acceptance—rejection rule. (Readers who wish to construct a ential scheme for variables may refer to the book edited by Dr L. Davies.[19]) Such a scheme will have its OC curve with its pro-er's and consumer's risks and risk points; and from an album of such mes a suitable one can be chosen. Usually, information is required ut the variability of the product. So far, however, nothing parallel he average outgoing quality limit has been proposed, and I am not re of any examples where the economically optimum degree of pling has been determined. It may be that too much in the problem pecific to the particular material under investigation to make ible the development of generalised schemes.

One development of general interest is epitomised in the British ndard 771 for synthetic resin (phenolic) moulding materials.[18] This dard specifies a number of tests to be made on the materials, and ic' values of the mean and standard deviation which a batch of the erial should have to conform to the specification. A batch is raised through a sample and the batch is regarded as conforming he specification provided that the sample value is better than a ain limit calculated from the basic value, the limit being so chosen if the batch has the basic value the probability of the sample value g better than the limiting value is 0·99. In other words, if the basic e is the producer's risk point, the producer's risk is 0·01; the sumer's risk and risk point are not brought into consideration. In respect the standard resembles the US Military Standard 105A.

even physical tests of an item of the material are of 'primary im-ance' and four are 'of interest to the user', and the appraisal is ensive. Consequently small samples of 2, 3, or 5 from a batch are gested; and these give the consumer little protection. This situation et by permitting the pooling of the results of a sequence of samples n from a sequence of batches produced over a period not exceeding months. With the larger pooled sample thus obtained the limits are e to the basic values and the protection to consumer is much roved. Again, as in Military Standard 105A, the effect of the dard is less to secure discrimination between individual satisfactory unsatisfactory batches than to appraise runs of batches from a given ducer. In this kind of development, acceptance sampling merges systematic quality control by the producer according to the ideas ciated with the control chart.

REFERENCES

[1] Dudding, B. P. (1955). *Drafting Specifications based on Limiting the Number of Defecti permitted in Small Samples.* B.S. 2635: 1955. British Standards Institutio London.

[2] Peach, Paul (1947). *Industrial Statistics and Quality Control.* 2nd edition. Edwa & Broughton, Raleigh NC.

[3] Horsnell, G. (1954). 'The determination of single-sample schemes for percenta defectives', *Appl. Statist.*, **3**, 150.

[4] Grubbs, F. E. (1949). 'On designing single sampling inspection plans', *Ann. Ma Statist.*, **20**, 242.

[5] Anscombe, F. J. (1949). 'Tables of sequential inspection schemes to cont fraction defective', *J. R. Statist. Soc.*, A, **112**, 180.

[6] Golub, A. (1953). 'Designing single-sampling inspection plans when the samp size is fixed', *J. Amer. Statist. Assoc.*, **48**, 278.

[7] Military Standard 105A (1950). *Sampling Procedures and Tables for Inspection Attributes.* Government Printing Office, Washington DC.

[8] Bellinson, H. R. (1950). Contribution to a discussion, pp. 45-51. *Accepta Sampling.* American Statistical Association, Washington DC.

[9] Statistical Research Group, Columbia University (1948). *Sampling Inspecti* McGraw-Hill, New York and London.

[10] Dodge, H. F. and Romig, H. G. (1945). *Sampling Inspection Tables.* Wiley, N York; Chapman & Hall, London.

[11] Greb, D. J. (1949). 'AOQL single sampling plans from a single chart and tab *J. Amer. Statist. Assoc.*, **44**, 62.

[12] Hamaker, H. C. (1951). 'Economic principles in industrial sampling proble a general introduction', *Bull. Int. Statist. Inst.*, **33**, V, 105.

[13] Anscombe, F. J. (1950). 'The cost of inspection', *Statistical Method in Indust Production*, pp. 61-67. Royal Statistical Society, London.

[14] Sittig, J. (1951). 'The economic choice of sampling system in acceptance sam ling', *Bull. Int. Statist. Inst.*, **33**, V, 51.

[15] Weibull, I. (1951). 'A method of determining inspection plans on an econo basis', *Bull. Int. Statist. Inst.* **33**, V, 85.

[16] Champernowne, D. G. (1953). 'The economics of sequential sampling procedu for defectives', *Appl. Statist.*, **2**, 118.

[17] Horsnell, G. (1957). 'Economical acceptance sampling schemes', *J. Roy. Sta Soc.*, A, **120**, 148.

[18] British Standard 771 (1948). *Synthetic Resin (Phenolic) Moulding Materials.* Bri Standards Institution, London.

[19] Davies, O. L., Editor (1957). *Statistical Methods in Research and Production,* 347-50. Oliver & Boyd, London and Edinburgh.

[20] Hamaker, H. C., Taudin Chabot, J. J. M. and Willemze, F. G. (1950). " practical application of sampling inspection plans and tables', *Ph Technical Review*, **11**, 362.

[21] Military Standard 414 (1957). *Sampling Procedures and Tables for Inspectio Variables for Per Cent Defective.* Government Printing Office, Washing DC.

Note added in Proof

In preparing this article I overlooked reference 20, which gi sampling tables based on the percentage defective at which the p bability of acceptance is 0·5. Reference 21, which has belatedly co to my notice, gives a method closely related to Military Standard 10

SOME BASIC PRINCIPLES OF SAMPLING INSPECTION BY ATTRIBUTES

HUGO C. HAMAKER

Philips Research Laboratories, Eindhoven, Netherlands

In this article Dr Hamaker discusses various approaches to one of the main problems in sampling inspection—that of deciding what size of sample to use.

oduction

ne of the chief problems in starting a sampling inspection procedure decide what size of sample is needed. The question can be solved arious ways, the most common being:

by applying one of the standard sampling inspection tables such as the Dodge and Romig Tables,[1] the Military Standard 105A,[2] or the Philips SSS Table[3];

by choosing numerical values for a suitable set of parameters (AOQL, Producer's or Consumer's Risk Point, etc.) and constructing a corresponding sampling plan; or

on the basis of an economic theory which takes into consideration various costs; see the literature cited in reference 4.

three methods have been applied with success. However, although hods (1) and (2) may be convenient, one can never feel quite ain that they will lead to what must be considered an optimum ple size. They may, as we shall see, lead in extreme cases to sample s many times larger than really necessary (see p. 157). The third hod suffers from the drawback that very often the basic costs under- g the theory cannot be estimated with a reasonable degree of cision.

he main purpose of this paper is to reconsider the fundamental blem of finding an optimum sample size in the light of recent papers he subject. I shall not attempt an exhaustive literature survey, but nly review what would seem to me to be the basic principles lved.

Purpose of Sampling Inspection

mpling inspection procedures may serve a dual purpose, namely:

) the discrimination of good lots from bad;

) to collect information on the quality of the lots inspected.

all briefly distinguish these as *discrimination* and *information*. Existing ries deal almost exclusively with the first of these problems, because

it lends itself more easily to theoretical treatment. There is consequentl some danger that we may lose sight of the second aspect, which neve theless may be the more important of the two.

If a consumer does not inspect the products received, the supplie may conclude that quality does not matter and he will relax in h efforts to maintain the quality at the best level possible. Hence routin inspection procedures may exert a decided influence on the quality the lots received and may be desirable for this reason alone, a vie which has also been expressed by I. D. Hill in the discussion of a pape by Horsnell.[4] In this connection an effective feed-back of informatio from inspector to producer or supplier may be of interest. Recer papers on vendor rating systems and their contribution to the improv ment of quality and of consumer-vendor relations[5] give evidence of development along these lines.

I know of only one paper, however, in which an attempt has bee made to incorporate into a theory the effect of sampling inspection o the quality of the lots received. Whittle[6] assumes that for a constar lot size the process average $\bar{p}$ of the lots received depends on th inspection sample size n by

$$\bar{p} = p_u + p_a e^{-cn} \qquad \ldots . (1$$

a function of the type depicted in Fig. 1;

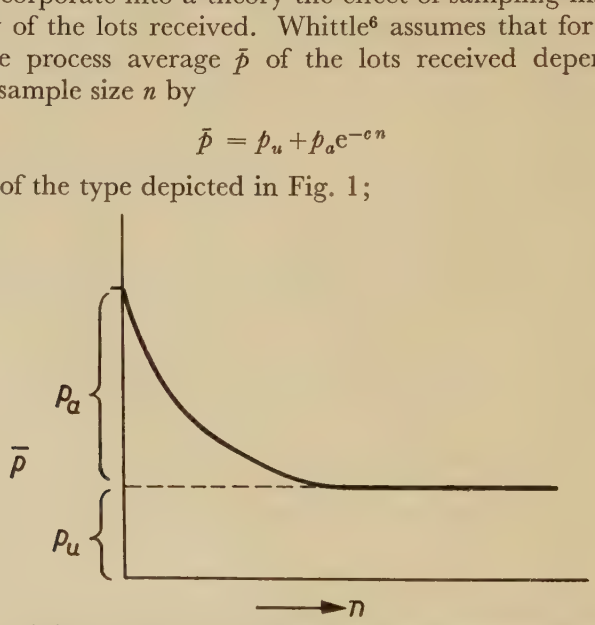


FIG. 1. The relation between process average $\bar{p}$ and sample size n according to Whittle.

p_u is the unavoidable percentage defective inherent in the productio process, p_a is the avoidable percentage, due to lack of attention, whi can be eliminated by exerting pressure through sampling inspectio c is a positive constant.

Whittle uses this formula to discuss the best allocation of a given tot of inspection efforts over a number of different products for each which a relation like (1) is known to hold good.

Relation (1) might also be introduced into a theory for finding t optimum sample size for a single product. If we did so, however, o concept of what is 'optimum' would have to be drastically revised. F

xisting theories assume a constant process average or a stationary ess curve, and this basic assumption would have to be abandoned. can easily imagine that in certain situations Whittle's principle ld justify the maintenance of a routine inspection by samples too ll to be effective in discriminating good lots from bad. Consequently nomic theories so far proposed should be looked upon with some icion; it is conceivable that they disregard what is in reality one of most important functions of inspection procedures: continually to ind the people interested that quality does matter. The same ciple is operative in, and I believe is mainly responsible for, the ess of control charts in improving production.

Process Curve

or convenience I shall restrict the following arguments to cases re rejected lots undergo 100% inspection.

$$\begin{aligned} p &= \text{fraction defective} \\ f(p) &= \text{process curve} = \text{distribution density of } p \\ &\quad \text{amongst lots submitted for inspection} \\ I &= \text{cost of inspecting one item} \\ D &= \text{loss caused by an accepted defective} \end{aligned} \qquad \ldots . \ (2)$$

the average we have to inspect $1/p$ items in order to detect one ctive; the cost involved is I/p. Hence there exists a critical or ak-even' fraction defective

$$p_c = I/D$$

e knew the fractions defective in the lots received, it would be most nomical to accept without inspection all lots for which $p<p_c$ and to % inspect all lots for which $p>p_c$; for $p=p_c$ both procedures would cceptable, the costs being the same.

nother important point is that in discriminating between good and lots sampling inspection is wholly ineffective when the process curve esents a binomial distribution; for then, by a theorem of Mood,[7] fraction defective in the not-inspected portions of the lots has just same distribution for accepted as for rejected lots. The lots are duced by a process which, in a statistical sense, is ideally stable and can no longer effectively distinguish between good and bad lots.

aken together these results lead to the conclusion that for separating d lots from bad, sampling inspection can only be effective if the ess curve shows a considerable spread across the break-even fraction ctive p_c. Hence, in setting up a sampling plan some information ut the process curve is important.

hough detailed information may frequently be missing, the simple stion: whether bad lots do occasionally cause trouble, may suffice practical purposes. If the answer is 'yes' sampling inspection is th trying, if 'no' it should be questioned whether sampling inspection eally needed. That is, if discrimination is the only problem; if

information is also important a routine inspection may still be justifiabl but then presents a different problem.

In a theoretical treatment of the discriminative operation of samplin schemes, however, the practical approach adopted above does n suffice; we have to introduce the process curve in the guise of som mathematical formula and more detailed assumptions concerning th shape of the distribution must be made.

In a recent paper, for example, Horsnell[4] assumes that what he cal '*normal*' production produces a binomial process curve. But, thoug this is not explicitly stated, it is evident from his article and the di cussion following it that in addition he assumes the occasional occu rence of '*bad*' lots with a much higher fraction defective. Thus h actual process curve is represented by a two-humped distribution, th humps lying widely apart (Fig. 2).

A different approach is that of 'Discovery Sampling', see Taylor Here a distinction is made between three mutually exclusive classes lots

I 100% effective lots	:	$p = 0$
II partially defective lots	:	$0<p<1$
III 100% defective lots	:	$p = 1$

Class III does not constitute a sampling risk because these lots are di covered by inspection of a few items only. Hence in setting up sampling scheme we have only to consider classes I and II. For the the process curve assumed in Discovery Sampling is

$$\begin{aligned} f(p) &= (1-A) \quad \text{for} \quad p = 0 \\ f(p) &= A(s+1)(1-p)^s \quad \text{for} \quad p>0 \end{aligned} \qquad \ldots\ldots ($$

That is, a fraction $(1-A)$ of the lots received is considered free fro defectives, the remaining fraction of partially defective lots having $(s+1)(1-p)^s$ distribution; A and s are constants to be estimated fro observation.

Other formulae that have been used for the process characteristic a $(s+1)(1-p)^s$ by Sittig,[9] $Bp^m(1-p)^n$ by Champernowne,[10] and tw fractions defective p_1 and p_2 occurring with frequencies f_1 and f_2 Barnard.[11] The characteristic shapes of these different process curv are indicated in Fig. 2.

Perhaps the distinction between the three classes made by Taylor somewhat too rigorous. But in practical situations it will often, believe, make sense to consider a classification of this kind; for examp

I′ highly effective lots,
II′ lots of doubtful quality,
III′ highly defective lots.

In a way the various process curves in Fig. 2 may be said to diff mainly in the degree of importance attached to these different class Horsnell mainly considers the classes I′ and III′, Discovery Sampli

egards class III′ and takes only I′ and II′ into account. Sittig's and ampernowne's curves seem to pay particular attention to inter-diate lots, ignoring the classes I′ and III′, whereas Barnard's model ght be adjusted so as to cover a variety of different situations. In one

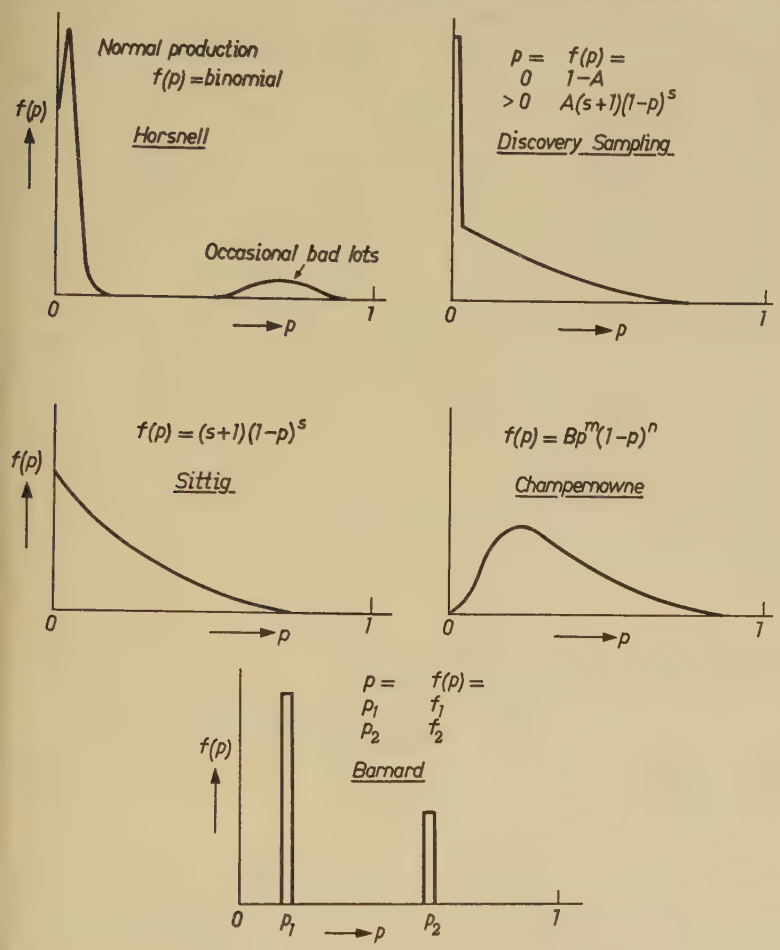


Fig. 2. The shapes of the process curve assumed by various authors.

ect all these process curves agree, namely, that there exists a con-rable variation in the fractions defective of the lots submitted for ection.

iscovery Sampling claims to have derived its process curve from nsive observations. The constant s in (3) has been found to be t 3 in many cases; the fraction A of partially defective lots has to etermined from sample data in any particular situation before overy Sampling can be applied.

his paper Horsnell[4] (Table I) sums up eight different equations he process characteristic, all in a general sense representing curves those of Sittig or Champernowne in Fig. 2. In the discussion to snell's paper Barnard observes: 'We at Imperial College did some

work in trying to find out what sort of process curves do in fact turn u in industry and none we have seen bears the slightest resemblance t those tabulated in Table I'.

Accumulated sampling inspection records might, of course, furnis some further information, but the samples taken are usually small an of varying size so that large and irregular sampling fluctuations woul have to be taken into account. Apart from this, industry is constantl changing its products and processes, and by the time we have collecte a sufficient number of data for a more detailed analysis some chang may be introduced which completely alter the situation. In fact i private correspondence Professor Barnard has mentioned that a Imperial College they found these difficulties insuperable and that n general pattern emerged.

It does not seem likely, therefore, that we shall ever possess ver precise information as to the shape of the process curve, and it als seems doubtful whether the process curve assumed in Discover Sampling can be accepted as a universal basis for the construction sampling procedures.

The Minimax Principle

The *minimax principle* has been applied to cope with our lack knowledge about the process curve. It can best be explained by a example. With

$$\begin{aligned} \mathcal{N} &= \text{lot size} \\ n &= \text{sample size} \\ P &= \text{probability of acceptance} \\ H &= \text{loss due to a rejected item} \end{aligned}$$

and the other symbols as in the previous section, Moriguti[12] defines th risk R per lot as

$$R = In + DNpP + HN(1-P) \qquad \ldots . \ (4$$

For any given sampling plan this risk will go through a maximum as function of the fraction defective p, and we choose our sampling pla so as to minimise this maximum.

Various criticisms can be raised against this procedure. First, it evident that the risk R is, under certain simplifying assumptions, th average cost per lot received if these all contain the same fraction d fective p. Since sampling inspection only makes sense if there is considerable spread in p, it seems unrealistic to consider this particul situation and use it as a basis for deciding the sample size. Furthermo the value of p corresponding to the minimax solution may well l outside the range of values of p which actually occur. The minima principle may perhaps be of value in an isolated case. But in industri situations we always possess some information concerning the proce curve; though this information may not be mathematically very exac yet it is not necessary to pretend complete ignorance.

econdly, the risk R measures the average cost *per lot inspected.* But in nomic problems we must not merely consider the cost, but the cost elation to what is achieved for it. We inspect material because we d it and the achievement is not measured by the lots inspected but by *lots accepted.* It would be more appropriate to consider the risk per epted lot, which will be $R/(1-P)$. This would completely alter the ure. In Moriguti's paper we find, for instance, that for the rule: ect all lots without inspection', the risk would be HN; per accepted the risk would be infinite. This may be an extreme example but statements do not help to convince us that the basic principles are nd. (Moriguti does not in fact apply the principle of minimax risk, a modification called the principle of minimax regret. The ob- ions raised hold equally well, however.)

third difficulty in applying equation (4) is that the costs D and annot in most practical situations be estimated with any reasonable cision, and may be extremely variable as well. The loss D due to epted defectives may encompass a loss of the consumers' good-will, assembly work it may depend on the quality of other components. loss H due to rejected items may vary with the stock available; if stock of some raw material is low, rejection of a lot may mean page of a production line, while if the stock is high rejection of a btful lot may not do much harm.

nomic Theories

ll economic theories of sampling inspection[13] start from (4) or e similar equation, which is then multiplied by the process curve and integrated from 0 to 1 to give the average cost

$$K = \int_0^1 Rf(p)\mathrm{d}p \qquad \ldots\ldots (5)$$

then have to find a sampling plan which minimises K.

arious expressions for R have been proposed and have been dis- ed in an earlier paper.[14] It should be pointed out that there is some erence of opinion and that the equation may depend on the situation saged. It is, for example, important how rejected lots are disposed whether they can be screened and corrected, have to be scrapped, an be sold for other purposes. Also, when producer and consumer perate closely it will be appropriate to minimise the overall cost, when such a close co-operation does not exist we may have to der the problem solely from the consumer's point of view. We not consider these problems in detail, but only discuss equation rom a general point of view.

hough we now no longer consider p as a constant, the other ob- ons raised against the minimax principle still hold good, while in tion we are faced with incomplete knowledge about the process e.

Horsnell[4] has tried to get away from this last difficulty by minimisin the average cost for his 'normal' production (Fig. 2) with a prescribe consumer's risk point that serves to account for the occasional 'bad lots. The main problem then is the choice of this risk point and th will depend on our concept of how bad the 'bad' lots are. If we fix th consumer's risk point arbitrarily, we are sub-optimising and we hav no assurance that the sample sizes obtained are anywhere near th true optimum.

General Conclusions

All in all I am inclined to conclude that the application of suc economic theories (minimax included) requires an amount of detaile information that is not as a rule available. They may be successful i isolated cases, but do not lead to simple principles with a wide field o application as primarily needed in industry.

In practice we should, therefore, always consider the setting up of sampling inspection procedure as a problem of trial and error. W start with a plan chosen as best we can on the information available but after a while we must collect the results and then decide whethe our first choice has been correct or whether changes are needed.

In a way this basic principle was already embodied in Dodge an Romig's original table, where they use the process average (obtaine from samples) as one of the parameters for choosing a sampling plan It is likewise incorporated in the Military Standard 105A, where th process average and other information supplied by the samples decide the transition from normal to tightened or reduced inspection, an vice versa.

Whether the process average is really adequate for this purpos would seem to me to be a debatable question. For in discriminatin between good and bad lots it is a large spread in the process curve tha makes sampling inspection really effective.

An interesting procedure (suggested to me by G. A. Barnard) migh be to inspect the samples in two halves, the numbers defective in eac half being separately recorded. We may then sentence a lot on th first half alone, and on the entire sample. Writing A for accepted, an R for rejected, the results could be recorded as AA, AR, RA, and RR If, after a while, we find very few AR's and RA's we may conclude tha the sample size can be halved. If these combinations occur too fre quently it could perhaps be concluded that the sample size should b increased. In such a manner we might sequentially arrive at a sati factory sample size.

An Interesting Case Study

Let us conclude by discussing an illustrative case study borrowe from Taylor.[8] It refers to a Company X producing nuts and bolts. I all 527 lots were inspected both by Military Standard 105A and b Discovery Sampling, the results being recorded in Table I.

TABLE I

Results of simultaneous inspection of 527 lots by Mil. Stand. 105A and by Discovery Sampling

		Discovery Sampling	
		Accepted	Rejected
Mil. Stand. 105A	Accepted	449	7
	Rejected	7	64
Average sample size for Mil. Stand. 105A : $\bar{n} = 86$ Disc. Sampling : $\bar{n} = 9$			
Average lot size : $\bar{N} = 3000$			

n 513 cases both plans led to the same decision, in only 14 they greed; this was despite the fact that the average number of obser-ons required for Discovery Sampling was only 10% of the number uired by Mil. Stand. 105A.

he results seem to prove that we have a process curve like that of snell in Fig. 2. Most lots are very good, some extremely bad, and a ll sample distinguishes between these two groups just as well as a e sample. Only a few lots are of intermediate quality and whether accept or reject these should not make much difference. Taylor ed mentions that of the 64 rejected lots some 38 gave 100% ctive samples. In applying Mil. Stand. 105A normal inspection an AQL of 0·65% was used.

iscovery Sampling always uses single samples of size n and an ptance number 0. The probability of acceptance then is

$$P = (1-p)^n \qquad \ldots . (6)$$

a process curve (3) the average outgoing quality becomes

$$\mathrm{AOQ} = \int_0^1 Pp\mathrm{f}(p)\mathrm{d}p = A\,\frac{(s+1)}{(s+n+1)(s+n+2)} \qquad \ldots . (7)$$

a function of s the AOQ goes through a maximum, the average oing quality limit, for which we find approximately

$$\mathrm{AOQL} = \frac{A}{4n} \qquad \ldots . (8)$$

e case of Table I an AOQL of 0·5% = 0·005 was prescribed and some preliminary observations it was concluded that A, the frac-of partially defective lots, was of the order of 0·20. We then obtain

$$n = \frac{A}{4\ \mathrm{AOQL}} = 10 \qquad \ldots . (9)$$

ylor gives separate results for nuts and for bolts, which required

different sample sizes. He does not state what value of A he actuall used. Also his formula for the AOQ is more complicated than (7 because he takes into account the fact that defective items are remove from the 100% inspected lots and the lot size is consequently reduced This is a minor correction, however, which does not alter the fina result (8).]

Thus the main argument behind Discovery Sampling is that we firs estimate from observations the fraction A of partially defective lots an then account for these by an AOQL based on the process curve (3) with s as an unknown parameter.

The question then arises: what does an example such as the one jus cited really prove? The situation is evidently an exceptional one, mos of the lots being either free of defectives or else very highly defective Also Discovery Sampling is a procedure which considers each case o its own merits, some information on the process curve being require before the system can be put into operation. It would be interesting t know how much preliminary sampling is needed before Discover Sampling can be applied. In comparing sample sizes the amount o this preliminary sampling should be taken into account.

Furthermore, it goes without saying that if we wish to keep the risk of wrong decisions low for the wide variety of products and circumstance for which Military Standard 105A was developed, we have on the aver age to take larger samples than when we investigate the requirement for each product separately, as Discovery Sampling does.

The only conclusion that can legitimately be drawn from Taylor' example, therefore, seems to be that blind adherence to a particula Sampling Table may occasionally lead to unnecessarily large sampl sizes. But this does not prove that the Table was wrong or that Dis covery Sampling is right. It would be desirable to compare bot systems in a large number of cases, including cases that are less ex ceptional than the one quoted as an example.

Conclusion

We wish to conclude by emphasising once more that pages 151-15 refer exclusively to the use of sampling inspection for discriminatin between good and bad lots. It is this problem which lends itself mo readily to theoretical discussion. There is some danger that we lo sight of the problem of information as outlined on pages 149-150 simpl because so much more has been written about the problem of di crimination.

It is not inconceivable that in actual practice the problem of info mation is at least as important and in the operation of samplin procedures this is a possibility which should be continually borne i mind.

REFERENCES

ODGE, H. F. and ROMIG, H. G. (1944). *Sampling Inspection Tables.* John Wiley, New York.

ILITARY STANDARD 105A (1950). *Sampling Procedures and Tables for Inspection by Attributes.* US Government Printing Office. Also reproduced in DUNCAN, A. J. (1952). *Quality Control and Industrial Statistics.* R. D. Irwin, Chicago.

AMAKER, H. C., TAUDIN CHABOT, J. J. M. and WILLEMZE, F. G. (1950). 'The practical application of sampling inspection plans and tables', *Philips Technical Review*, **11**, 362.

ORSNELL, G. (1957). 'Economical acceptance sampling plans', *J. R. Statist. Soc.*, A, **120**, 148.

CQUEMARD, F. C. (1957). 'Start quality control at the source', *Mill and Factory*, **60**, 104.

IZGIBBONS, R. G. (1955). 'The Bendix Radio vendor quality rating system', *Industr. Qual. Contr.*, **11**, No. 8, 38.

EORGIO, G. (1957). 'An application of the Bendix quality rating system to a small plant', *Industr. Qual. Contr.*, **13**, No. 9, 12.

HITTLE, P. (1954). 'Optimum preventative sampling', *J. Operat. Res. Soc. Amer.*, **2**, 197.

OOD, A. M. (1943). 'On the dependence of sampling inspection plans upon population distributions', *Ann. Math. Statist.*, **14**, 415.

AYLOR, E. F. (1957). 'Discovery sampling by attributes', *National Convention Transactions of the ASQC*, 517.

TTIG, J. (1951). 'The economic choice of sampling systems in acceptance sampling', *Bull. Int. Statist. Inst.*, **33**, V, 51.

HAMPERNOWNE, D. G. (1953). 'The economics of sequential sampling procedures for defectives', *Appl. Statist.*, **2**, 118.

RNARD, G. A. (1954). 'Sampling inspection and statistical decisions', *J. R. Statist. Soc.*, B, **16**, 151.

ORIGUTI, S. (1955). 'Notes on sampling inspection plans', *Reports of Statistical Applications Research, Union of Japanese Scientists and Engineers*, **3**, No. 4, 1.

e the list of literature cited in reference 4.

AMAKER, H. C. (1951). 'Economic principles in industrial sampling problems', *Bull. Int. Statist. Inst.*, **33**, V, 105.

THE LOGNORMAL DISTRIBUTION IN QUALITY CONTRO

JAMES MORRISON

Standard Telephones and Cables Limited, Paignton, Devon

In the application of statistical control chart analysis t thermionic valve test data, difficulties have been encountere because of the skewness of the parent distribution. Mr Morriso discusses the generation of these skew distributions in theoretic and practical terms, and describes a modified quality contr scheme for these cases; ratio is used instead of range as measure of variability, and the logarithmic transformation used to calculate control limits.

Introduction

The well-established position which statistical quality control enjo in present-day industry must be a great source of satisfaction to th pioneers in this branch of applied statistics. The success of the standar control techniques is due in large measure to the fact that a great man industrial processes generate data which is normally distributed (o nearly so) and which is therefore amenable to treatment by relativel simple processes of arithmetic and graphical analysis. The apparen universality of the normal distribution tends to obscure the fact tha skew distributions can occur quite naturally in some industrial process and do not necessarily signify lack of control.

It is not surprising that warning signals will appear as soon as ske distributed data is fed into the standard control routine, for this wa designed specifically for detecting departures of any sort from a pre cribed normal distribution. It is important to recognise (from exper ence with the process or from technical knowledge of the product) whe the skew distribution is the correct pattern. It is then necessary to tak steps to restore the precision and sensitivity of the control chart analys so that it will prove just as reliable with skewed data as it would hav been with normal data.

In a recent paper Moore[1] discusses the effect of various forms non-normality and gives a warning against too rigid an interpretatio of the control limits. Prior to this the subject of skew distributions i quality control had not received a thorough systematic treatmen Earlier papers[2, 3, 4] in which basic principles were laid down conta only passing reference to skew distributions. Hill's modified limit provide a higher margin of safety when manufacturing to a specific tion, but being arranged symmetrically they are not ideally suited f skew distributions. Moshman's critical values[6] are useful for tests significance when the universe parameters are known, but they are n sufficient for calculating control limits when only a few samples a

lable. The logarithmic transformation is familiar in other branches atistical analysis, but its use in quality control does not appear to e been exploited. Reference should be made to the monograph by hison and Brown[7] for the basic properties of the lognormal distri- on which have been used in setting up the lognormal quality rol system.

the radio valve industry it has been found that much of the test is skewed, some of it to a very marked degree. Experience with ibutions such as those in Fig. 1 has indicated that the problem is of

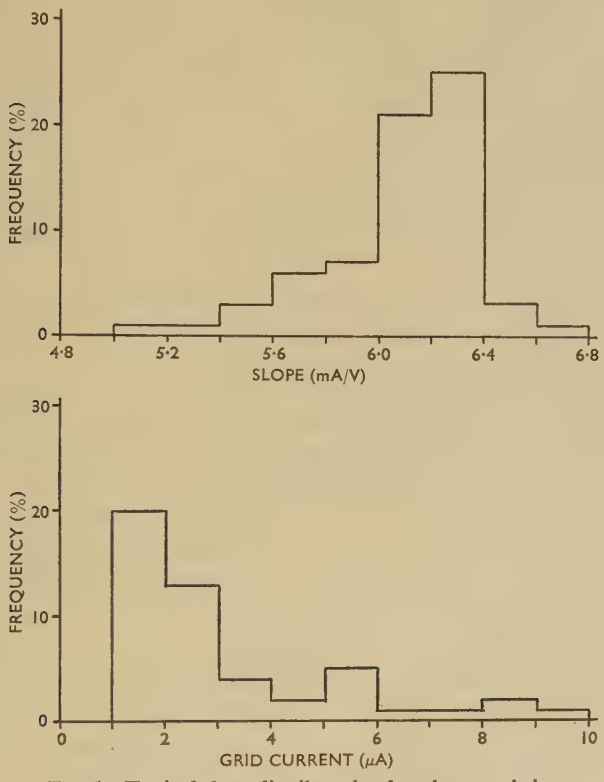


Fig. 1. Typical skew-distributed valve characteristics.

ient practical importance to justify the development of the special rol chart routine described later in this paper. This routine is based ne properties of the lognormal distribution, which in the present of statistical theory is the easiest of all skew unimodal distributions nipulate.

s of Skew Distributed Test Data

is interesting to reflect on the reason why valve test data should imes be skew distributed. In some circumstances data which are ally distributed when under control can assume the appearance skew distribution when control breaks down. This may be due to resence of outliers beyond the usual range of the distribution, and

to one side of it. It does not follow that a skew distribution alway signifies lack of control in the system which generates it. As will b shown later there are circumstances in which skewness must be regarde as being typical of a variate. The question whether a skew distributio is a natural occurrence, or whether it signifies lack of control, is on which can only be resolved by practical experience or by a study of th system which generates the variate in question.

A thermionic valve can be regarded as a complex physical syste whose measurable characteristics can be classified as dependent variate (i.e. test data) which are a function of independent primary variate (e.g. electrode dimensions, emissivity, temperature, degree of vacuum etc.). Leaving aside the possibility that the primary variates may b skew distributed, and assuming that they are normally distributed, on can show on general theoretical grounds that skew distributions ar to be expected in the dependent variates.

Gibrat's theory of the law of proportionate effect[8, 9] provides on explanation of the genesis of skew distributions in valve technology. statement of this law (due to Aitchison and Brown) is as follows: ' variate subject to a process of change is said to obey the law of propo tionate effect if the change in the variate at any step of the process is random proportion of the previous value of the variate.' It can b argued that many of the processes of valve manufacture will obey th law. For example, during activation the cathode is heated unde vacuum until a suitable state of electron emission is reached. Th change in emissivity during processing depends on various factors (suc as heater wattage) which may be randomly distributed, and so th process can be held to obey the law of proportionate effect. It can b shown mathematically that under certain conditions the law results i the generation of lognormal variates[10] and so it will not be surprising t find skew distributions among valve characteristics.

An alternative approach is provided by Haldane,[11] who shows tha the powers and products of normal variates are in general skew distr buted and leptokurtic. Haldane also shows that on certain assumptio the nth power of a normal variate tends to a lognormal variate as increases.

By applying electric field theory to certain idealised forms of th thermionic valve it is possible to establish mathematical relations b tween the electrical behaviour and the valve geometry. For exampl the amplification factor of a low-mu plane triode is given by Spange berg[12] as:

$$\mu = \frac{-2\pi d_{gp}}{a \log\left(2 \sin \frac{\pi r_g}{a}\right)}$$

where d_{gp} = distance between grid and anode planes
a = grid-wire spacing
r_g = grid-wire radius.

is and other formulae employ powers and products of the primary tes, as well as trigonometric and exponential functions. One can fore expect, from Haldane's result, that some of the dependent tes (i.e. the electrical characteristics) will be skew distributed, and l be reasonable to attempt to fit a lognormal distribution to such

me of the known skew distributions of valve characteristics can be ined in more practical terms. Two examples will be discussed. e are: (*a*) the measurement of the cut-off current; and (*b*) the current test used as a measurement of the goodness of emission.
the cut-off test the anode current is measured with sufficient grid applied to reduce the anode current (ideally) to zero. In practice all anode current is measurable even on good valves. Various ts, such as distorted grids, or cathode coating approaching too to the ends of grids, will always tend to *increase* the current. The nt of the increase develops rapidly with the degree of the fault, consequently the upper tail of the distribution is usually extended ve a positive skew.
the total current test the valve is set to run at stated conditions a specified value of cathode current. The cathode temperature is reduced and a new value of cathode current is measured. Good s will show only a small drop in current, whereas poor valves show a large drop. Here again the effect increases rapidly with egree of fault, and consequently the lower tail of the distribution al current is usually extended to give a negative skew.

Logarithmic Transformation

plication to Quality Control. The usual method for handling skew buted data in statistical analysis is to apply a logarithmic trans- ation before proceeding with the analysis. This approach would erfectly correct in control chart work, but it would not be very actory in industrial applications, where the task of recording the and plotting the charts is usually delegated to shop floor operators e mathematical ability may not be sufficient to deal with loga- ns. Moreover, it is easier for the technicians whose task is to inter- the charts, if these are plotted in the original scale of measurement, ne sample values can then be related directly to other data in their day experience (e.g. specification limits).
ere are strong arguments in favour of a control scheme which will s skew data in the original scale of measurement, using only tforward arithmetical operations. It will not be a serious dis- ntage if the calculation of control limits is a little more complex, this is generally the task of a competent supervisor.
the proposed scheme the fundamental relationships between al and lognormal distributions are used to derive control limits for ognormal variate from the corresponding limits for a normal te. This amounts to using the logarithmic transformation in re-

verse to fix the control limits, so that the sample values can be plot in their original scale of measurement.

Hill[5] has pointed out that the modified limits used for high-precis data are less robust than natural limits towards departures from n mality. It is worth noting that the lognormal scheme tackles t problem at its source, by making the correct basic assumptions for sk distributed data.

Types of Lognormal Distribution. Aitchison and Brown[7] discuss f forms of the lognormal distribution:

(*a*) The two-parameter distribution $\Lambda(\mu, \sigma^2)$ which describes posit skew data with a lower threshold of zero.
(*b*) The three-parameter distribution $\Lambda(\tau, \mu, \sigma^2)$ which describes p tive skew data with a lower threshold of τ. This can be reduced the two-parameter case by the transposition $X' = X - \tau$.
(*c*) The three-parameter distribution $\Lambda(\theta, \mu, \sigma^2)$ which describes ne tive skew data with an upper threshold of θ. This can be redu to the two-parameter case by the transposition $X' = \theta - X$.
(*d*) The four-parameter distribution $\Lambda(\tau, \theta, \mu, \sigma^2)$ which descri skew data with upper and lower thresholds of θ and τ.

For quality control purposes it will be sufficient to consider how control limits should be determined for the simple two-parameter tribution. The three-parameter distributions, which will occur m generally in practice, can be reduced to the two-parameter case applying the appropriate transposition. The four-parameter dis bution is relatively intractable, and is not likely to find practical ap cation in quality control work.

Ratio as a Measure of Variability. An important consequence of normal—lognormal relation is that ratio (r) is the appropriate meas of variability in skew samples, corresponding to range (w) in nor samples. For if we write $y = \log x$, and the values are ordered ascending magnitude from 1 to n, then

$$\begin{aligned} w &= y_n - y_1 \\ &= \log x_n - \log x_1 \\ &= \log (x_n / x_1) \\ &= \log r \end{aligned}$$

At first sight there might appear to be serious theoretical objecti to using ratio as a measure of variability, because of the large stand error; but the derivation of the standard error of the ratio pres difficulties, and in any case the figure is not of much use owing to skewness of the distribution of the ratio.

Fortunately, the large standard error of the ratio does not affect probability levels of the ratio control limits in lognormal quality cont which are simply a transformation of range control limits. If a sam ratio is outside the control limits in the lognormal scheme, then

sponding sample range in the transformed variate must also be de the range control limits. The interpretation of ratio and range ts is therefore exactly the same whether one uses the lognormal m to analyse the skew data, or whether one uses the normal control m to analyse the transformed values of the data.

ometric Mean or Arithmetic Mean? Strictly speaking, the geometric ns of samples should be used in the lognormal system in order to rve the correspondence with the arithmetic mean of the normal m, but this would detract from the essential simplicity of the pro- re. In practice the errors due to the use of the arithmetic mean are troublesome in very skew distributions extending over two or more s of the logarithmic scale. In such cases the sensitivity of the control s will be affected, but this can be allowed for by recalculating the netric mean for those borderline samples (relatively few in number) e arithmetic means fall just beyond a limit.

here is, of course, no objection to the use of the geometric sample n in circumstances where high precision is important (e.g. in ytical work in a laboratory). The labour of calculating geometric ns can be reduced a little by constructing a nomogram for the *n*th of the product of the sample values.

ne possibility of using a sample median value, arrived at by arrang- he individuals in order of magnitude, has been considered, but this not appear to be satisfactory. In an experiment with 120 random les of five drawn from a lognormal distribution it was found that entral limit of the median values was in close agreement with that e geometric means, but the distribution of median values extended a much greater range than either the arithmetic or the geometric ns, especially in the direction of the long tail of the parent distri- n. The median value of the sample cannot therefore be regarded sufficiently accurate measure of its mean value.

ntrol Limits. The control limits for a normal distribution are cal- ed by processes of addition, subtraction, multiplication, and divi- The corresponding processes for a lognormal distribution will be iplication, division, evolution, and involution. Hence the ratio of le values in a lognormal distribution corresponds to the range n samples in a normal distribution, and so the factors used as ipliers of the average range in calculating normal distribution ol limits should be used as indices of the average ratio in the rmal distribution. (Strictly speaking, the geometric mean ratio to be used, but the arithmetic mean ratio is a sufficiently good ximation.)

e lognormal scheme can be linked with any of the existing versions ality control which are based on an assumed normal distribution. various control limit factors which are tabulated in the literature ences 2, 3, 4, 5, and 13) will continue to be used, but the formulae lculating the limits will be rather different. A summary of the ols and definitions used in lognormal distributions is given in

Table I, and in Tables II and III the lognormal control limit formul are compared with their normal counterparts.

TABLE I

Comparison of normal and lognormal formulae (distributions and parameters)

	Normal	Lognormal
Distributions	$X = \mathrm{N}(\mu, \sigma^2)$	Two-parameter, positive skew: $X = \Lambda(\mu, \sigma^2)$ Three-parameter, positive skew: $X = \Lambda(\tau, \mu, \sigma^2)$ Three-parameter, negative skew: $X = \Lambda(\theta, \mu, \sigma^2)$
Parameters	μ = mean $= \frac{1}{n}\Sigma x$ σ^2 = variance $= \frac{1}{n-1}\Sigma(x-\bar{x})^2$	μ = mean $= \frac{1}{n}\Sigma \log x$ σ^2 = variance $= \frac{1}{n-1}\Sigma(\log x - \overline{\log x})^2$ τ = lower threshold θ = upper threshold
Transformations		Three-parameter, positive skew: $X' = X - \tau$ Three-parameter, negative skew: $X' = \theta - X$

NOTE: Data from a three-parameter lognormal distribution must *always* be reduced to the two-parameter case before calculating ratios and limits.

TABLE II

Comparison of normal and lognormal formulae (samples, specifications, and precision rating)

	Normal	Lognormal
Sample values:		
Individuals (in ascending order of magnitude)	$x_1, x_2, x_3 \ldots . x_n$	$x_1, x_2, x_3 \ldots . x_n$
Sample mean	$\bar{x} = \frac{1}{n}\Sigma x$	$\bar{x} = \frac{1}{n}\Sigma x$
Sample variability	$w = x_n - x_1$	$r = x_n/x_1$
Average variability (over k samples)	$\bar{w} = \frac{1}{k}\Sigma w$	$\bar{r} = \frac{1}{k}\Sigma r$
Specification limits (see note):		
Maximum limit	$x_{\max}$	$x_{\max}$
Minimum limit	$x_{\min}$	$x_{\min}$
Tolerance	$w_s = x_{\max} - x_{\min}$	$r_s = x_{\max}/x_{\min}$
Precision rating:		
Relative precision index	R.P.I. $= w_s/\bar{w}$	R.P.I. $= \frac{\log r_s}{\log \bar{r}}$
High precision	R.P.I. $> 3{\cdot}5$	R.P.I. $= 3{\cdot}5$
Medium precision	$3{\cdot}5 \geqq$ R.P.I. $\geqq 2{\cdot}5$	$3{\cdot}5 \geqq$ R.P.I. $\geqq 2{\cdot}5$
Low precision	$2{\cdot}5 >$ R.P.I.	$2{\cdot}5 >$ R.P.I.

NOTE: In a skew distribution, one of the specification limits may lie beyond the threshold, clear of the distribution. Relative to the threshold, this limit will have a negative value, and the formula for R.P.I. given above will be meaningless. In such cases the data should be placed in the high-precision class.

TABLE III

Comparison of normal and lognormal formulae (control limits)

	Normal	Lognormal
rol limit factors:		
ample average (*B.S.* 600R)	A'	A'
Iodified (*B.S.* 2564)* ..	A''	A''
ange (*B.S.* 600R)	D'	D'
rol limits:		
ample mean target	$x_t = \frac{1}{2}(x_{\min} + x_{\max})$	$x_t = \sqrt{(x_{\min} \times x_{\max})}$
ample mean upper limit ..	$x_t + A'\bar{w}$	$x_t \tilde{r}^{A'}$
ample mean lower limit ..	$x_t - A'\bar{w}$	$x_t \tilde{r}^{-A'}$
Iodified limits for Upper	$x_{\max} - A''\bar{w}$	$x_{\max} \tilde{r}^{-A''}$
igh precision data Lower	$x_{\min} + A''\bar{w}$	$x_{\min} \tilde{r}^{A''}$
ample variability Range	$D'\bar{w}$	—
upper limit Ratio	—	$\tilde{r}^{D'}$

NOTE: The formulae for lognormal distributions apply only to the two-parameter distribution (See Table I).

* Readers who prefer to do so could use the factor $\mathcal{J}$ given by Hill in *Applied Statistics*, **5**, 12.

reshold Values (τ and θ). The quality control scheme which is sum-ed in the tables is only suitable for the simple two-parameter bution with positive skew and zero threshold. In practical cases l be necessary to determine the threshold value (τ or θ), which is in one of the transpositions ($X' = X - \tau$) or ($X' = \theta - X$) to reduce ata to a simple two-parameter distribution.

some cases the threshold can be determined by reference to the ating system. In the example of the total current test discussed r the initial value of the current can be regarded as the threshold, in the cut-off example discussed previously a zero threshold afely be assumed. In other cases the threshold might be estimated an inspection of the frequency distribution of the data.

reshold values can be determined accurately by calculation, but omputation is rather involved. The graphical method described itchison and Brown[7] will probably be sufficiently accurate for ty control purposes. This consists of plotting on logarithmic ability paper the cumulative distribution curves for several feasible s t of τ, and of selecting the value of t which gives the nearest pach to a straight line.

cal Application of Lognormal Quality Control

essor Ackoff[14] has shown that the efficiency of operational solu-an be improved by using control procedures to detect changes in stem parameters. In another field Dr Box[15] has demonstrated the iple of evolutionary operation of industrial processes. Both these iples can be embodied with advantage in quality control, so con-g a routine control procedure into a more active self-correcting n which will not only survive, but which will develop and thrive, ficult circumstances.

The simplest way in which to adopt the necessary 'control with control' strategy is to treat the record sheet (on which the sample da are recorded) as being equal in importance to the control chart. ringing the out-of-limit entries on the record sheet in a distincti

TABLE IV

Example of record sheet, showing 'normal' treatment of skewed data

Sample No.		1	2	3	4	5	6	7	8	9	10
Data:	1	4·55	1·93	2·22	2·71	2·91	2·20	2·82	2·76	4·98	4·8
	2	4·99	3·95	1·73	2·45	5·68	5·66	5·22	4·40	4·05	2·7
	3	3·62	4·10	5·10	4·60	4·33	3·71	3·75	3·13	4·00	3·5
	4	3·52	4·16	4·52	2·09	3·51	3·35	3·50	1·55	7·20	3·1
	5	3·77	1·61	4·06	1·90	3·24	1·61	3·31	3·70	3·18	4·8
Total		20·45	15·75	17·63	13·75	19·67	16·53	18·60	15·54	23·41	19·0
Mean		4·090	3·150	3·526	2·750	3·934	3·306	3·720	3·108	4·682	3·8
Range		1·47	2·55	3·37	2·70	2·77	4·05	2·40	2·85	4·02	2·1

Sample No.		11	12	13	14	15	16	17	18	19	20
Data:	1	4·50	3·07	2·39	2·92	2·56	2·53	3·41	2·85	3·16	2·5
	2	1·95	4·02	2·91	4·25	4·38	4·16	3·10	4·16	3·70	4·7
	3	3·41	4·17	3·09	3·02	1·24	3·78	6·02	3·17	2·61	1·6
	4	2·87	4·33	3·15	2·26	2·62	3·77	1·09	2·50	2·65	2·6
	5	1·90	4·06	2·52	5·72	1·92	1·72	2·92	3·91	3·42	3·5
Total		14·63	19·65	14·06	18·17	12·72	15·96	16·54	16·59	15·54	15·1
Mean		2·926	3·930	2·812	3·634	2·544	3·192	3·308	3·318	3·108	3·0
Range		2·60	1·26	0·76	3·46	3·14	2·44	4·93	1·66	1·09	3·1

Sample No.		21	22	23	24	25	26	27	28	29	30
Data:	1	3·61	3·16	2·96	2·47	3·55	1·35	2·30	3·71	9·48	1·9
	2	2·13	4·20	6·09	3·49	3·35	2·50	2·26	3·06	1·72	2·5
	3	5·08	2·32	3·78	3·38	3·18	2·51	2·22	1·53	4·20	4·2
	4	2·01	2·44	2·29	4·45	4·75	4·20	1·60	2·45	3·37	3·1
	5	1·92	1·62	4·16	2·61	8·72	3·50	9·70	6·40	5·58	1·9
Total		14·75	13·74	19·28	16·40	23·55	14·06	18·08	17·15	24·35	13·8
Mean		2·950	2·748	3·856	3·280	4·710	2·812	3·616	3·430	4·870	2·7
Range		3·16	2·58	3·80	1·98	5·54	2·85	8·10	4·87	7·76	2·3

'NORMAL' TREATMENT: Specification limits = 1·0/10·0

$\therefore$ Target = $\frac{1}{2}(1{\cdot}0 + 10{\cdot}0) = 5{\cdot}5$

Mean range = $(1{\cdot}47 + 2{\cdot}55 + 3{\cdot}37 + \ldots)/30 = 3{\cdot}20$

Upper control limit (mean) = $5{\cdot}5 + (0{\cdot}377 \times 3{\cdot}20) = 6{\cdot}71$

Lower control limit (mean) = $5{\cdot}5 - (0{\cdot}377 \times 3{\cdot}20) = 4{\cdot}29$

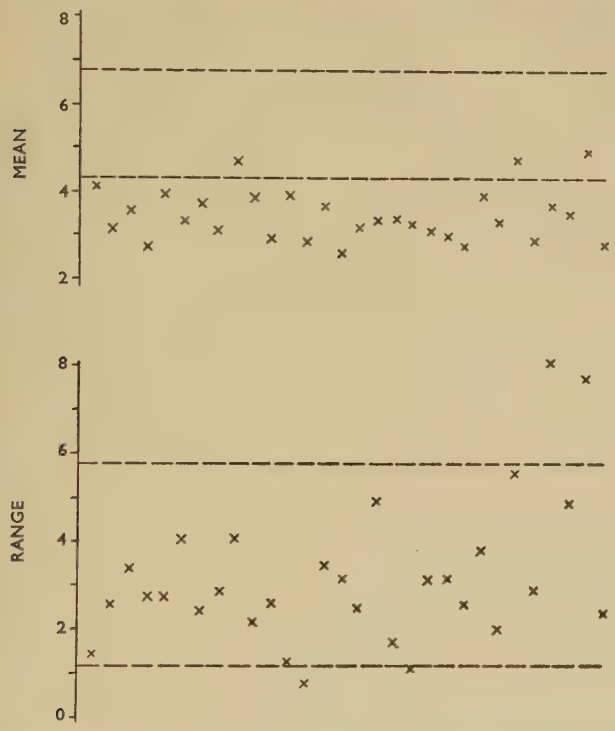
Upper control limit (range) = $1{\cdot}81 \times 3{\cdot}20 = 5{\cdot}79$

Lower control limit (range) = $0{\cdot}37 \times 3{\cdot}20 = 1{\cdot}18$

colour one can see at a glance whether the recorded data support diagnosis of the control charts. If the two are incompatible an inve gation should reveal some defect in the control system. It was adoption of this simple cross-check as a regular part of quality con routine which exposed the fallacious assumption of normality in cert

e characteristics and which led to the development of the lognormal ity control system.

the valve industry great importance is attached to the product ing certain specified test limits, and so the control charts are used arily to detect the risk of failure of individual valves (known quially as 'shrinkage'). The target for sample means should there- be determined from the specification limits, irrespective of what prevailing level of sample means may be (e.g. if the data are ally distributed the arithmetic mean of the limits should be taken e target). This is a departure from the usual practice of setting the



Control charts for moderately skewed data ('normal' treatment with upper and lower 1 in 40 limits).

ol limits on either side of the estimated mean of the samples, but easoning is perfectly sound. There is no point in having control at urrent mean level, if this level is one which produces shrinkage!

t us consider the problem of setting up new control charts for a hetical variate whose statistical properties are unknown at the , but which ultimately proves to be skew distributed. Table IV how the record sheet might appear for the first thirty samples of duct made to specified limits of 1·0 to 10·0. For simplicity's sake mal distribution would be assumed initially, and the control limits be calculated for medium precision data, with the target set at iddle of the specified range, so that the control charts would r as in Fig. 2. These charts depict a process running below target,

with a slight tendency to excessive variability, but reference back to th record sheet does not support this diagnosis, since none of the individua are outside the specification limits, as they would be in these circum stances if the distribution were normal.

At this point it would be sensible to stop and examine the shape the distribution. The cumulative distribution plotted on a logarithmi probability grid in Fig. 3 shows clearly that a lognormal distributio

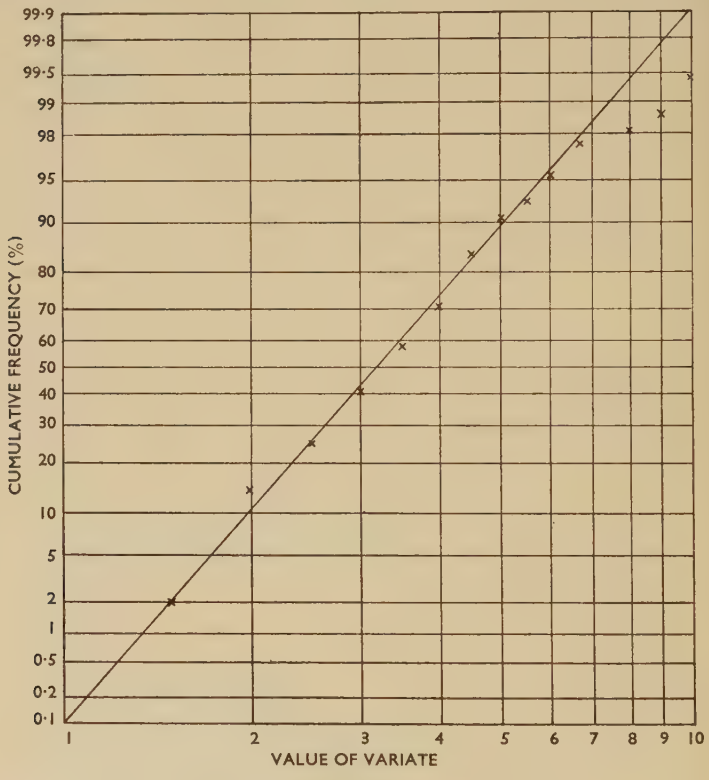


Fig. 3. Cumulative distribution of data in Table IV.

curve can be fitted quite accurately, and so in this case normal quali control should be replaced by lognormal quality control. This wou be done by substituting 'ratio' for 'range' in the record sheet and b plotting a ratio chart. The lognormal limits for the sample mean a 4·63 and 2·16 and it can easily be seen that these would span most the points on Fig. 2 quite well. The upper limit for the ratio is 6·21 a all the sample ratios would be found to be less than this.

Rather than show the charts for the lognormal transformation Fig. 2 in detail, I have chosen another example to show how the schem functions for a distribution of even greater skewness. Fig. 4 sho control charts for 120 random samples of five, drawn from the lo normal population ($\tau=0$, $\mu=0$, $\sigma=0{\cdot}322$) whose three-sigma limi embrace two cycles of values (i.e. from 0·1 to 10·0). The individu

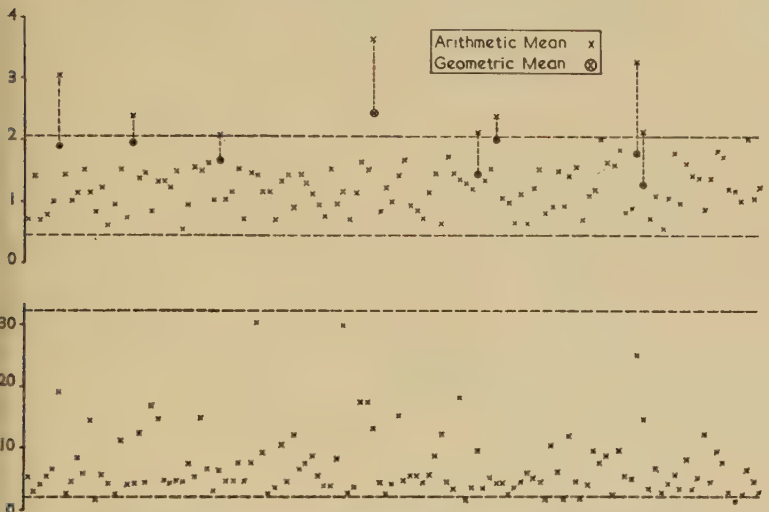
es (not given here) ranged from 0·10 to 9·30, with a median of 1·00. rol limits (1 in 40) were calculated as follows:

Target value $= x_t = \sqrt{(0{\cdot}1 \times 10{\cdot}0)} = 1{\cdot}0$
Mean ratio $= \bar{r} = 6{\cdot}851$

ntrol factors for sample of five:

$A' = 0{\cdot}377 \quad D'_L = 0{\cdot}37 \quad D'_U = 1{\cdot}81$

Sample mean upper limit	$= x_t \bar{r}^{A'}$	$= 2{\cdot}068$
Sample mean lower limit	$= x_t \bar{r}^{-A'}$	$= 0{\cdot}484$
Sample ratio upper limit	$= \bar{r}^{D'_U}$	$= 32{\cdot}4$
Sample ratio lower limit	$= \bar{r}^{D'_L}$	$= 2{\cdot}04$



Control charts for skewed data (lognormal treatment with upper and lower 1 in 40 limits).

e control charts in Fig. 4 show that the only control limit which nsgressed with excessive frequency is the upper control limit for le means, above which there are eight values instead of the ex- d three. By the use of the geometric means for these samples, this ber is reduced to one, which is rather less than the 'correct' ency. Nevertheless, most of the adjusted points lie close to the , and it seems reasonably certain that a trend away from the target, wards greater variability, would be detected quite readily. This ple shows that the charts work well for a very skew distribution, though the arithmetic mean ratio is employed in calculating ol limits, and the arithmetic sample mean is plotted and only ced by the geometric mean in borderline cases.

sion

e modified quality control scheme has been used with considerable s on skew distributed valve test data, and has led to a reasonably

accurate interpretation of the data where the standard quality contr routine had failed. The only practical problem which has arisen is th of deciding when to use the standard scheme, and when to use t modified scheme. The answer is to use the standard scheme whereve possible, and to confirm by reference to the original data that a dangerous trend which is signalled on the control charts does actuall exist. If it does not, the shape of the distribution should be studied a if it is found to be skew the lognormal control scheme should be adopte in place of the normal scheme.

REFERENCES

[1] Moore, P. G. (1957). 'Normality in quality control charts', *Appl. Statist.*, **6**, 17

[2] Pearson, E. S. (1935). *B.S.* 600: *The Application of Statistical Methods to Industri Standardisation and Quality Control*, pp. 130-131. British Standards Institutio London.

[3] Dudding, B. P. and Jennett, W. J. (1942). *B.S.* 600R: *Quality Control Charts*, p 25, 28, and 29. British Standards Institution, London.

[4] Dudding, B. P. and Jennett, W. J. (1955). *B.S.* 2564: *Control Chart Technique wh Manufacturing to a Specification.* British Standards Institution, London.

[5] Hill, D. (1956). 'Modified control limits', *Appl. Statist.*, **5**, 12.

[6] Moshman, J. (1953). 'Critical values of the lognormal distribution', *J. Am Statist. Ass.*, **48**, 600.

[7] Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution.* Cambrid University Press.

[8] Gibrat, R. (1930). 'Une loi des répartitions économiques: l'effet proportionn *Bull. Statist. Gén. Fr.*, **19**, 469.

[9] Gibrat, R. (1931). *Les Inégalités Économiques.* Libraire du Recueil Sirey, Paris.

[10] Aitchison, J. and Brown, J. A. C., *ibid.*, pp. 22-26.

[11] Haldane, J. B. S. (1942). 'Moments of the distributions of powers and produ of normal variates', *Biometrika*, **32**, 226.

[12] Spangenberg, K. R. (1948). *Vacuum Tubes*, p. 125. McGraw-Hill, New York.

[13] Freund, R. A. (1957). 'Acceptance control charts', *Industr. Qual. Contr.*, **14**, No. pp. 13-22.

[14] Ackoff, R. L. (1957). 'The concept and exercise of control in operations researc *Proceedings of the International Conference on Operational Research*, p. 26. Engli Universities Press, London.

[15] Box, G. E. P. (1957). 'Evolutionary operation', *Appl. Statist.*, **6**, 81.

NALYSIS OF VARIANCE OF A RANDOMISED BLOCK DESIGN WITH MISSING OBSERVATIONS*

WILLIAM ALEXANDER GLENN and CLYDE YOUNG KRAMER
Virginia Agricultural Experiment Station of the Virginia Polytechnic Institute

In this article the authors describe a general method for estimating several missing values and analysing the results so obtained. They give explicit formulae for all cases when two or three values are missing, formulae for some other specific cases, and an illustrative example.

duction

experimental work it sometimes happens that the results of one nore observations are lost. Animals may become sick or die, field may be ravaged by some pest, or labels may be mislaid. Despite me care in conducting the experiment, when the data are sum- sed there may be blank cells indicating missing observations.

hen the original design is accidentally upset, the methods of ysis which are ordinarily applicable must be modified. If there are ng values, the orthogonality property of the design disappears. procedure of the analysis of variance applicable to orthogonal riments cannot be used directly on non-orthogonal data. The tivity of the sums of squares computed in the usual fashion does nold.

lan and Wishart[1] developed a method of analysis for a randomised design with one missing observation. They derived a formula he missing value by minimising the error mean square and then eeded to analyse the augmented data in the usual manner except the degrees of freedom for error and total were each reduced by Yates[5] extended their method to cover several missing plots, the nates being obtained by an iterative procedure. He also showed in a complete analysis of variance using estimated values for the ng plots, the treatment sum of squares is overestimated but may be cted by subtracting the bias, for which he gave a generalised rla.

en[2] presented formulae to estimate two and three missing values 'id not attempt to correct for the bias in the treatment sum of es. His results appear to be more complicated than need be se he did not attempt any simplifications. Pires and Baptista[3]

condensation of a thesis by W. A. Glenn written under the direction of C. Y. Kramer mitted to the Virginia Polytechnic Institute in partial fulfilment of the requirements M.S. degree in Statistics.

considered three special cases but did not attempt to generalise th problem of missing values. Thompson[4] presented a method of analys for two special cases: (i) when no two observations are missing in th same treatment; and (ii) when no two observations are missing in th same block. Computationally his results are more complex than th results obtained in this paper and are special cases of our general cas

This paper will deal with the estimation of several missing values i the randomised block design, by minimising the error sum of square Explicit equations for each value missing will be derived for man special cases and will include in simplest form all the results of th above authors. A procedure will be given for the completely genera case which may prove to be less tedious in application than the iterativ method of Yates. A direct method of analysis of variance, not requirin a correction for bias in the treatment sum of squares, will be demon strated. Formulae will be given for the bias introduced if an analys of variance is carried out on the augmented data; these, thoug equivalent to that of Yates, differ from the latter slightly in form an are easier to use.

Model and Notation

It is assumed that an observation may be represented by the mathe matical model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij};\ i = 1, \ldots, p;\ j = 1, \ldots, q \quad \ldots . (2.1$$

where μ, τ_i, and β_j are constants such that $\sum_{i=1}^{p} \tau_i = \sum_{j=1}^{q} \beta_j = 0$;

μ represents the overall mean effect, τ_i represents the true effect of th ith treatment measured as a deviation from the mean, and β_j represent the true effect of the jth block measured as a deviation from the mean The ϵ_{ij}, representing random error effect, are independently and nor mally distributed with mean 0 and variance σ^2.

We shall denote treatment totals and block totals by

$$Y_{i.} = \sum_{j=1}^{q} y_{ij} \quad \text{and} \quad Y_{.j} = \sum_{i=1}^{p} y_{ij} \qquad \ldots . (2.2$$

if there are no missing values in the ith treatment and jth block, and l

$$\begin{aligned} Y'_{i.} &= \text{total of observed values in treatment } i \\ Y'_{.j} &= \text{total of observed values in block } j \end{aligned} \qquad \ldots . (2.3$$

if there are missing values in the ith treatment and jth block. We sha also let

$$Y'_{..} = \text{total of all observed values} \qquad \ldots . (2.4$$

nates of Missing Values

se of n Missing Values, No Two in the Same Block or Treatment. Let missing values be represented by x_{cd}, x_{ef}, x_{gu}, . . ., n in all, where wo of the first subscripts are equal and no two of the second subts are equal. By partially differentiating the expression for the sum of squares with respect to the x's and equating the derivatives ro the following system of equations is obtained:

$$\begin{array}{lllll} -1)(q-1)+ & x_{ef} & + & x_{gu} & +\ldots = pY'_{c.}+qY'_{.d}-Y'_{..} \\ +x_{ef}(p-1)(q-1)+ & & & x_{gu} & +\ldots = pY'_{e.}+qY'_{.f}-Y'_{..} \\ + & x_{ef} & +x_{gu}(p-1)(q-1)+ & & \ldots = pY'_{g.}+qY'_{.u}-Y'_{..} \\ + & \ldots & + & \ldots & +\ldots = \quad \ldots \end{array} \qquad \ldots . \quad (3.1)$$

ning the quantity Z as follows:

$$Z_{ij} = pY'_{i.}+qY'_{.j}-Y'_{..} \qquad \ldots . \quad (3.2)$$

may express the above system of equations in matrix notation as

$$\begin{bmatrix} -1)(q-1) & 1 & . & 1 \\ 1 & (p-1)(q-1) & . & 1 \\ . & . & . & . \\ 1 & 1 & . & (p-1)(q-1) \end{bmatrix} \begin{bmatrix} x_{cd} \\ x_{ef} \\ \ldots \\ \ldots \end{bmatrix} = \begin{bmatrix} Z_{cd} \\ Z_{ef} \\ \ldots \\ \ldots \end{bmatrix} \qquad \ldots . \quad (3.3)$$

symmetric $n \times n$ matrix appearing in (3.3) is seen to have $(p-1)$) as each diagonal element and unity elsewhere. Its inverse is ly found to be the symmetric $n \times n$ matrix such that each diagonal ent is

$$\frac{(p-1)(q-1)+(n-2)}{(p-1)^2(q-1)^2+(n-2)(p-1)(q-1)-(n-1)}$$

very other element is

$$\frac{-1}{(p-1)^2(q-1)^2+(n-2)(p-1)(q-1)-(n-1)}$$

e solution of (3.3) is thus given by

$$x_{hk} = \frac{[(p-1)(q-1)+(n-2)]Z_{hk} - \sum_{\substack{\text{missing plots} \\ i \neq h \quad j \neq k}} Z_{ij}}{(p-1)^2(q-1)^2+(n-2)(p-1)(q-1)-(n-1)} \qquad \ldots . \quad (3.4)$$

e h, k take on successively the pairs of values corresponding to ng plots. The summation in the numerator refers to the sum of Z s for all missing plots except hk.

e of n Missing Values, All in a Single Block or Treatment. Consider he case in which n values are missing in treatment r. Let the missalues be represented by x_{rc}, x_{rd}, x_{re}, ... As in the previous n a system of equations is obtained by partially differentiating

the expression for the error sum of squares. This system is expressed i matrix notation as

$$\begin{bmatrix} (p-1)(q-1) & (1-p) & . & (1-p) \\ (1-p) & (p-1)(q-1) & . & (1-p) \\ . & . & . & . \\ (1-p) & (1-p) & . & (p-1)(q-1) \end{bmatrix} \begin{bmatrix} x_{rc} \\ x_{rd} \\ \cdots \\ \cdots \end{bmatrix} = \begin{bmatrix} Z_{ro} \\ Z_{rd} \\ \cdots \\ \cdots \end{bmatrix} \quad \ldots . \text{(3.5}$$

The symmetric $n \times n$ matrix appearing in (3.5) has $(p-1)(q-1)$ a each diagonal element and $(1-p)$ elsewhere. Its inverse is readil found to be the symmetric $n \times n$ matrix such that each diagonal el ment is

$$\frac{q-n+1}{(p-1)[(q-1)^2-(n-2)(q-1)-(n-1)]}$$

and every other element is

$$\frac{1}{(p-1)[(q-1)^2-(n-2)(q-1)-(n-1)]}$$

The formula for the estimation of n missing values, all in treatment (where $n<q$) is

$$x_{rk} = \frac{(q-n+1)Z_{rk} + \sum_{\substack{\text{missing plots} \\ j \neq k}} Z_{rj}}{(p-1)[(q-1)^2-(n-2)(q-1)-(n-1)]} \quad \ldots . \text{(3.6}$$

where k takes on successively the values corresponding to missing plo in treatment r. The summation in the numerator refers to the sum Z values for all missing plots in treatment r other than rk.

In the same manner one obtains the formula for the estimation n missing values, all in block s (where $n<p$) as

$$x_{hs} = \frac{(p-n+1)Z_{hs} + \sum_{\substack{\text{missing plots} \\ i \neq h}} Z_{is}}{(q-1)[(p-1)^2-(n-2)(p-1)-(n-1)]} \quad \ldots . \text{(3.}$$

where h takes on successively the values corresponding to missing plo in block s. The summation in the numerator refers to the sum of values for all missing plots in block s other than hs. In the process deriving (3.7), it is found that the matrix in the equation analogous (3.5) differs from the matrix in (3.5) only in the fact that p and q a interchanged.

Estimation of Two Missing Values. If only two values are missing, the must necessarily fall into one of the three categories discussed in th two previous sections. The appropriate formulae are obtained fro (3.4), (3.6), and (3.7).

If two values are missing, not in the same block or treatment, l them be represented by x_{cd} and x_{ef}, $c \neq e$ and $d \neq f$, then

$$x_{cd} = \frac{(p-1)(q-1)Z_{cd} - Z_{ef}}{(p-1)^2(q-1)^2-1} \qquad x_{ef} = \frac{(p-1)(q-1)Z_{ef} - Z_{cd}}{(p-1)^2(q-1)^2-1} \quad \ldots . \text{(3.}$$

two values are missing, both in treatment r, let them be repre-
d by x_{rc} and x_{rd}. Then (3.6) gives

$$= \frac{(q-1)Z_{rc}+Z_{rd}}{(p-1)[(q-1)^2-1]} \qquad x_{rd} = \frac{(q-1)Z_{rd}+Z_{rc}}{(p-1)[(q-1)^2-1]} \quad \ldots . \quad (3.9)$$

two values are missing, both in block s, let them be represented by
nd x_{ds}. Then (3.7) gives

$$= \frac{(p-1)Z_{cs}+Z_{ds}}{(q-1)[(p-1)^2-1]} \qquad x_{ds} = \frac{(p-1)Z_{ds}+Z_{cs}}{(q-1)[(p-1)^2-1]} \quad \ldots . \quad (3.10)$$

timation of Three Missing Values. When three values are missing in
ndomised block design, they must occur in one of the following
ngements:

se 1: No two in the same block or treatment;
se 2: All three in the same block or treatment;
se 3: Value rs missing together with another value from treatment r, and another from block s;
se 4: Two values missing in treatment r, together with a third neither in treatment r nor in the same block as either of the two;
se 5: Two values missing in block s, together with a third neither in block s nor in the same treatment as either of the two.

ses 1 and 2 are covered by formulae (3.4), (3.6), and (3.7) with

r case 3 let the missing values be represented by x_{rc}, x_{rs}, and x_{ds}
e $d \neq r$ and $c \neq s$. The resulting equations, after differentiating,
en in matrix notation are

$$\begin{bmatrix} -1)(q-1) & (1-p) & 1 \\ 1-p) & (p-1)(q-1) & (1-q) \\ 1 & (1-q) & (p-1)(q-1) \end{bmatrix} \begin{bmatrix} x_{rc} \\ x_{rs} \\ x_{ds} \end{bmatrix} = \begin{bmatrix} Z_{rc} \\ Z_{rs} \\ Z_{ds} \end{bmatrix} \quad (3.11)$$

ll be convenient to introduce the following notation:

$$v = 1-p, \quad w = 1-q \qquad \ldots . \quad (3.12)$$

rms of which (3.11) assumes the form

$$\begin{bmatrix} vw & v & 1 \\ v & vw & w \\ 1 & w & vw \end{bmatrix} \begin{bmatrix} x_{rc} \\ x_{rs} \\ x_{ds} \end{bmatrix} = \begin{bmatrix} Z_{rc} \\ Z_{rs} \\ Z_{ds} \end{bmatrix} \quad \ldots . \quad (3.13)$$

olution of (3.13) is thus given by

$$x_{rc} = \frac{wZ_{rc}-Z_{rs}}{v(w^2-1)} \qquad \ldots . \quad (3.14)$$

$$x_{rs} = \frac{(v^2w^2-1)Z_{rs}}{vw(v^2-1)(w^2-1)} - \frac{Z_{rc}}{v(w^2-1)} - \frac{Z_{ds}}{w(v^2-1)} \quad \ldots . \quad (3.15)$$

$$x_{ds} = \frac{vZ_{ds}-Z_{rs}}{w(v^2-1)} \qquad \ldots . \quad (3.16)$$

For case 4 let the missing values be x_{rc}, x_{rd}, and x_{ef} where $e \neq r$ an c, d, f are all different. For this case we must solve

$$\begin{bmatrix} vw & v & 1 \\ v & vw & 1 \\ 1 & 1 & vw \end{bmatrix} \begin{bmatrix} x_{rc} \\ x_{rd} \\ x_{ef} \end{bmatrix} = \begin{bmatrix} Z_{rc} \\ Z_{rd} \\ Z_{ef} \end{bmatrix} \quad \ldots . \text{(3.17}$$

The solution of (3.17) is given by

$$x_{rc} = \frac{1}{v^2w^2 + v^2w - 2}\left\{\frac{v^2w^2 - 1}{v(w-1)} Z_{rc} + \frac{1 - v^2w}{v(w-1)} Z_{rd} - Z_{ef}\right\} \quad \ldots . \text{(3.18}$$

$$x_{rd} = \frac{1}{v^2w^2 + v^2w - 2}\left\{\frac{v^2w^2 - 1}{v(w-1)} Z_{rd} + \frac{1 - v^2w}{v(w-1)} Z_{rc} - Z_{ef}\right\} \quad \ldots . \text{(3.19}$$

$$x_{ef} = \frac{1}{v^2w^2 + v^2w - 2}\left\{v(w+1) Z_{ef} - (Z_{rc} + Z_{rd}) \right. \quad \ldots . \text{(3.20}$$

For case 5 let the missing values be represented by x_{cs}, x_{ds}, and $x_{}$ where $f \neq s$ and c, d, e are all different. We must now solve

$$\begin{bmatrix} vw & w & 1 \\ w & vw & 1 \\ 1 & 1 & vw \end{bmatrix} \begin{bmatrix} x_{cs} \\ x_{ds} \\ x_{ef} \end{bmatrix} = \begin{bmatrix} Z_{cs} \\ Z_{ds} \\ Z_{ef} \end{bmatrix} \quad \ldots . \text{(3.21}$$

The solution of (3.21) is given by

$$x_{cs} = \frac{1}{w^2v^2 + w^2v - 2}\left\{\frac{w^2v^2 - 1}{w(v-1)} Z_{cs} + \frac{1 - w^2v}{w(v-1)} Z_{ds} - Z_{ef}\right\} \quad \ldots . \text{(3.22}$$

$$x_{ds} = \frac{1}{w^2v^2 + w^2v - 2}\left\{\frac{w^2v^2 - 1}{w(v-1)} Z_{cs} + \frac{1 - w^2v}{w(v-1)} Z_{ds} - Z_{ef}\right\} \quad \ldots . \text{(3.23}$$

$$x_{ef} = \frac{1}{w^2v^2 + w^2v - 2}\left\{w(v+1) Z_{ef} - (Z_{cs} + Z_{ds})\right\} \quad \ldots . \text{(3.24}$$

General Case. When n missing values in a randomised block desig are distributed arbitrarily over the blocks and treatments, it is no possible to give an explicit equation for each value missing, coverin all permissible values of n. However, it is possible to set down certai features which apply in general.

When the criterion of minimising the error sum of squares is applied the unknowns (x's) representing missing values are obtained as th solutions of a system of linear equations which may be expressed i matrix notation as

$$\mathbf{UX} = \mathbf{Z} \quad \ldots . \text{(3.25}$$

where $\mathbf{U}$ is an $n \times n$ symmetric matrix to be described below, $\mathbf{X}$ is a $n \times 1$ column vector of unknowns representing missing values, and $\mathbf{Z}$ an $n \times 1$ column vector of corresponding Z values as defined by (3.2). I

ases considered above the solution was effected by inverting the atrix to give a result of the form

$$\mathbf{X} = \mathbf{U}^{-1}\mathbf{Z} \quad \ldots . \quad (3.26)$$

e particular structure of the **U** matrix depends on the manner in n the missing values are distributed relative to one another over locks and treatments. To facilitate discussion of this, consider the of n missing values and let the vacant plots be numbered in any from 1 to n. Let the unknown representing the ith missing value is numbering system be x_i. This x_i plays the same part as the ous x_{ij}, of course, but the single-subscript designation is more enient for present purposes. The $n \times 1$ column vector **X** is thus up of elements x_i, where i takes on integral values from 1 to n. makes possible a complete description of the $n \times n$ matrix **U**, n is always symmetric. The nature of the equations from which ults indicates that

) every diagonal element is $(p-1)(q-1)$;
) if x_i and x_j are not in the same block or treatment, element ij is 1;
) if x_i and x_j are in the same treatment, element ij is $(1-p)$; and
) if x_i and x_j are in the same block, element ij is $(1-q)$.

ng written the column vector **X**, one may construct the matrix **U** the above properties by referring to the data table. The column r **Z**, consisting of elements determined from (3.2) for each vacant must be written in the order which corresponds to that of the x_i's. system (3.25) is thus obtained.

sis of Variance

Direct Method of Analysis. This method avoids the direct calculation e bias in the treatment sum of squares arising from analysis of the ented data (i.e. including estimates of missing values) by the lard method for a randomised block design. The estimates for the ng yields are first entered in the table. The analysis of the aug- ed data is then carried out as usual. However, in recording the of squares for total and blocks (before correction for the mean) a ation into two parts is made for later reference. In the case of sum of squares one part corresponds to values originally present, ther part to values inserted. In the case of blocks one part gives um of squares for complete blocks, while the other part arises from s in which values were inserted. No separation is necessary in the ent sum of squares. A test of significance of treatment effect may ade on the augmented data in the regular manner, the degrees of om for each of total and error being reduced by one for each ng value estimated. Should this approximate test prove non- icant, one may stop there without fear of having drawn a wrong usion. This is because the bias always enhances the apparent icance.

If, however, the approximate test indicates significance, one proce to find the correct treatment sum of squares to be tested against er This is accomplished by deleting the inserted values and treating data as a one-way classification with unequal numbers in the blo The work is facilitated by referring to the separation in sums of squa mentioned above. The total sum of squares (before correction for mean) has been calculated as the part corresponding to values orig ally present. The block sum of squares is found by adding to the va already obtained for complete blocks the part corresponding to blo having missing plots. Note must be taken of the actual number in block in performing the usual divisions. The appropriate correct factor for the mean based on data originally present is then subtrac from each of the above, giving quantities T' and B'. The cor treatment sum of squares (A') is found by subtracting from the t sum of squares for the one-way classification (T') the block sum squares for the one-way classification (B') and the error sum of squa obtained from the augmented data (E^*). This method will be il trated on p. 184.

Analysis of Variance by Correcting for the Bias. Formulae for the bias the treatment sum of squares may be obtained by generalising procedure of the previous section. The theory on which this is ba has been given by Yates.[5] Let T^*, B^*, A^*, and E^* refer to sums squares (corrected for the mean) associated with total, blocks, tre ments, and error respectively when the standard method of analysi applied to the augmented data. Let T', B', A', and E' refer to corresponding quantities when the correct least-squares procedur carried out on the known data. In terms of this notation we may Yates's theory to write

$$E^* = E' \quad \ldots . (4$$

and

$$A' = (B' + A') - B' \quad \ldots . (4$$

where

$$(B' + A') = T' - E' = T' - E^* \quad \ldots . (4$$

Since we may calculate E^* from the augmented data as

$$E^* = T^* - B^* - A^* \quad \ldots . (4$$

we may write (4.3) as

$$B' + A' = T' - T^* + B^* + A^* \quad \ldots . (4$$

where T' is the total sum of squares of the original yields with missing values (less correction for the mean). Now B' may be obtai from the original block totals as indicated in the previous section. T we have from (4.5)

$$A' = T' - T^* + B^* + A^* - B'$$

or

$$\text{Bias} = A^* - A' = T^* - T' + B' - B^* \quad \ldots . (4$$

since the bias is the amount by which A^* exceeds A'.

using (4.6) to express the bias in terms of quantities in the data it is convenient to distinguish between the following two cases: hen no two of the missing values are in the same block, and (ii) more than one value is missing in a given block.

SE (i). Suppose that n values are missing, no two being in the block. Let x_{ij} represent estimates for these, where i, j take on the s corresponding to the missing plots. Then one has

$$= \sum\sum_{\substack{known\\data}} y_{ij}^2 + \sum\sum_{\substack{missing\\plots}} x_{ij}^2 - \frac{\left\{Y'_{..} + \sum\sum_{\substack{missing\\plots}} x_{ij}\right\}^2}{pq}$$

$$= \sum\sum_{\substack{known\\data}} y_{ij}^2 - \frac{Y'^2_{..}}{pq-n}$$

$$= \sum_{\substack{known\\data}} \frac{Y^2_{.j}}{p} + \sum_{\substack{missing\\plots}} \frac{Y'^2_{.j}}{p-1} - \frac{Y'^2_{..}}{pq-n}$$

$$= \sum_{\substack{known\\data}} \frac{Y^2_{.j}}{p} + \sum\sum_{\substack{missing\\plots}} \frac{(Y'_{.j}+x_{ij})^2}{p} - \frac{\left\{Y'_{..} + \sum\sum_{\substack{missing\\plots}} x_{ij}\right\}^2}{pq}$$

bining these according to (4.6), and simplifying the result, one ns

$$\text{Bias} = \frac{1}{p(p-1)} \sum\sum_{\substack{missing\\plots}} \{Y'_{.j} - (p-1)x_{ij}\}^2 \qquad \ldots\ldots (4.7)$$

SE (ii). For the case in which n values are missing, all in the kth , the derivation is similar to the above. The result obtained is

$$= \sum_{\substack{missing\\plots}} x^2_{ik} + \frac{Y'^2_{.k}}{p-n} - \frac{\left\{Y'_{.k} + \sum_{\substack{missing\\plots}} x_{ik}\right\}^2}{p} \qquad \ldots\ldots (4.8)$$

is clear that (4.8) holds only if $n<p$. If n were equal to p the kth block would be missing.

determine the bias in a general case one should proceed as follows: (4.7) sum over all blocks with one missing value; apply (4.8) $=2$ to all blocks having two missing values; apply (4.8) with o all blocks having three missing values, and so on. The total bias sum of the terms thus obtained.

er the total bias has been obtained, the correct treatment sum of s (A') is calculated by subtracting the total bias from the treat- sum of squares (A^*) calculated in the usual manner from the ented data.

Example

Illustrations, given below, of the application of the techniq described above are based on a randomised block experiment w three missing values. The experiment was concerned with the inve gation of phosphorus fertilisation for alfalfa. The yields are presen in Table I.

TABLE I

Total annual yield of alfalfa in pounds of airdried hay

	Blocks						Original Totals	Augmented Totals
	1	2	3	4	5	6		
Treatments: 1	13·75	18·30	22·39	23·72	21·86	23·60	123·62	123·62
2	13·05	22·69	21·73	24·29	23·27	23·23	128·26	128·26
3	18·67	17·51	18·38	22·83	23·84	23·61	124·84	124·84
4	15·36	17·45	18·39	20·97	21·31	23·86	117·34	117·34
5	*18·44*	22·16	23·42	*25·50*	20·75	28·52	94·85	138·79
6	21·87	21·34	23·73	*26·18*	27·84	21·93	116·71	142·89
Original Totals ..	82·70	119·45	128·04	91·81	138·87	144·75	705·62	
Augmented Totals ..	101·14	119·45	128·04	143·49	138·87	144·75		775·74

NOTE: Estimates of missing values are in italics and are underlined for emphasis.

e values underlined in Table I have been estimated. It is readily that this example belongs to *Case* 3 on p. 177. To obtain the ates we first substitute the appropriate original totals in (3.2) to n

$$Z_{51} = 6(94{\cdot}85) + 6(82{\cdot}70) - 705{\cdot}62 = 359{\cdot}68$$

imilarly

$$Z_{54} = 414{\cdot}34 \text{ and } Z_{64} = 545{\cdot}50$$

from (3.12), (3.14), (3.15), and (3.16)

$$x_{51} = \frac{-5(359{\cdot}68) - 414{\cdot}34}{-5(24)} = 18{\cdot}44$$

$$x_{54} = \frac{(624)(414{\cdot}34)}{25(24)(24)} + \frac{359{\cdot}68}{120} + \frac{545{\cdot}50}{120} = 25{\cdot}50$$

$$x_{64} = \frac{-5(545{\cdot}50) - 414{\cdot}34}{-5(24)} = 26{\cdot}18$$

estimated values are then inserted in the data table and the ented totals are obtained.

xt we compute T^*, B^*, A^*, and E^* by using the method on p. 179.

$$\frac{T_{..}^2}{36} = \frac{(775{\cdot}74)^2}{36} = 16\,715{\cdot}9041$$

$$T^* = 15\,469{\cdot}2556 + 1\,675{\cdot}6760 - 16\,715{\cdot}9041 = 429{\cdot}0275$$

$$B^* = \frac{70\,899{\cdot}9835}{6} + \frac{30\,818{\cdot}6797}{6} - 16\,715{\cdot}9041 = 237{\cdot}2064$$

$$A^* = \frac{100\,766{\cdot}4494}{6} - 16\,715{\cdot}9041 = 78{\cdot}5041$$

$$E^* = T^* - B^* - A^* = 113{\cdot}3170$$

e approximate test is set up as shown in Table II. Since three s have been inserted, the degrees of freedom for total and error each been reduced by three.

TABLE II

Approximate analysis of variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	*F*
ks	5	237·2064	47·4413	
tments ..	5	78·5041	15·7008	3·05
r	22	113·3170	5·1508	
l	32	429·0275		

Since the approximate test indicates significance at the five per ce level, the appropriate tabular value of F being 2·66, we proceed find the correct treatment sum of squares to be tested against erro Using the original totals and treating the data as a one-way classif cation with unequal numbers in the blocks we calculate

$$\frac{Y'^{2}_{..}}{33} = \frac{(705{\cdot}62)^2}{33} = 15\,087{\cdot}8662$$

$$T' = 15\,469{\cdot}2556 - 15\,087{\cdot}8662 = 381{\cdot}3894$$

$$B' = \frac{70\,899{\cdot}9835}{6} + \frac{6\,839{\cdot}2900}{5} + \frac{8\,429{\cdot}0761}{4} - 15\,087{\cdot}8662 = 203{\cdot}924$$

$$A' = T' - B' - E^* = 64{\cdot}1477$$

In the above it may be noted that in getting T' use was made 15 469·2556 previously calculated in connection with T^*. Similar 70 899·9835, the sum of squares related to complete blocks used getting B^*, appeared also in B'. The analysis is completed by settir up the exact analysis of variance appearing in Table III.

TABLE III

Exact analysis of variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Blocks	5	203·9247	40·7849	
Treatments ..	5	64·1477	12·8295	2·49
Error	22	113·3170	5·1508	
Total	32	381·3894		

When the above calculated F is compared with 2·66, the appropria tabular value at the five per cent level, it is concluded that the effe due to treatments is hardly significant. This means that the bias this case is sufficient to modify the conclusion from significance in t approximate test to doubtful significance in the exact test.

The above method of analysis avoids the direct use of a formula f the bias. However, should one wish to calculate the bias separately, may be obtained by the use of formulae (4.7) and (4.8). Use of (4. for block 1 yields

$$\text{Bias} = \frac{\{82{\cdot}70 - 5(18{\cdot}44)\}^2}{6(5)} = 3{\cdot}0083$$

and use of (4.8) with $n=2$ for block 4 yields

$$\text{Bias} = (25{\cdot}50)^2 + (26{\cdot}18)^2 + \frac{(91{\cdot}81)^2}{4} - \frac{(143{\cdot}49)^2}{6} = 11{\cdot}3480$$

Taking the sum of these gives the total bias as 14·3563. It is noted th this agrees with the difference between the augmented and corre

ment sums of squares, apart from a rounding error of one unit in ourth decimal place.

REFERENCES

AN, F. E. and WISHART, J. (1930). 'A method of estimating the yield of a missing plot in field experimental work', *J. Agric. Sci.*, **20**, 399.

EN, W. D. (1939). *Formulas for finding Estimates for Two and Three Missing Plots in Randomized Block Layouts.* Tech. Bull. 165, Michigan State College.

ES, A. J. and BAPTISTA, J. G. (1955). 'Analise do delineamento em blocos casualizados quandon existem dois e tres talhoes perdidos', *Melhoramento*, **8**, 61.

OMPSON, H. R. (1956). 'Extensions to missing plot techniques', *Biometrics*, **12**, 241.

ES, F. (1933). 'The analysis of replicated experiments when the field results are incomplete', *Emp. J. Exp. Agric.* **1**, 129.

PROFESSOR HOGBEN'S 'CRISIS'—A SURVEY OF THE FOUNDATIONS OF STATISTICS

D. V. LINDLEY
Statistical Laboratory, Cambridge

In this article Mr Lindley combines a review of a recent bo by Professor Hogben with an essay on some aspects of t fundamental theory of statistics. He deals in particular wi the work of Professor Savage and its implications.

THE subject-matter of statistics has points of contact with almost eve branch of science, and accordingly it is not surprising to find th several of the more important contributors to the advancement of t subject are scientists whose main field of interest lies elsewhere. Pr fessor Hogben is a prominent member of this circle of contributors, ar after publishing his very large book in two volumes, *Chance and Cho by Cardpack and Chessboard*, he now offers us a third[1] which, althoug not so large, is still substantial. In it he attempts an 'overdue revalu tion of statistical theory'. He has a subtitle 'The Relationship of Prob bility, Credibility and Error', and a sub-subtitle 'An Examination the Contemporary Crisis in Statistical Theory from a Behaviouri Viewpoint'. The news that there is a crisis in statistics at the mome will come as a surprise to most statisticians, but will undoubted stimulate interest in the book, especially amongst those who have r high regard for the subject anyhow.

Because of the prestige of the author and the importance of the thes many statisticians will want to know something about the book. I length and the difficulty of the English style in which it is couched m deter many from reading it at first hand, and therefore a fairly extensi summary seems desirable here. A foreword explains the behaviouri viewpoint that is the basis of the book: a viewpoint which concentrat on our manner of knowing and our methods of communication, rath than on a static collection of knowledge. The first chapter is entitl 'The Contemporary Crisis or the Uncertainties of Uncertain Inferenc It begins with an interesting division of the Theory of Statistics in four topics: (i) the calculus of errors, in which we seek for best appro mations to a value whose measurements are disturbed by error; (ii) t calculus of aggregates, embracing kinetic theory and Mendelism; (i the calculus of exploration, a topic which is not easy to define b embraces regression and factor analysis; (iv) the calculus of judgemen The chapter concludes with a justification of the use of the word 'cris and quotes several statisticians at length for this purpose. Now it certainly true that statisticians are interested in the foundations

subject and like to discuss them, and thereby provide a source for quotations, but does this imply the existence of a crisis? Any ect in which such discussion does not take place is either dead, or o young to have foundations. Statistics has only recently ceased g too young and the foundations are very ripe for exploration. author's argument here does not seem strong enough to establish se of a strong word like 'crisis', though, as we shall see when we lown to details, he is right in thinking that perhaps some revalua- of statistical thinking is necessary.

e book proper is divided into four parts. The first part is called Founding Fathers' and contains five chapters. The first is con- d with gambling. Probability theory began with the practical lem of the division of stakes in a gambling game, and Hogben s its early development through a definition in terms of favourable unfavourable cases to discussions involving frequencies. Proba- es resident in the mind are dismissed as having no frequency inter- tion. In the second chapter the relevance of randomness in ying probability laws is explored, the works of Venn and von Mises g mentioned. The connection with insurance is discussed in the chapter; the central limit theorem is explained and no less than neanings of frequency distribution are listed (p. 105).

is is perhaps the place to comment on the author's curious English It suffers from three serious defects: lack of punctuation, over lication, and inadequate precision. Consider, for example, the ving sentence to illustrate the first point. 'The most we can say t the relevance of the classical theory to the task of the Actuary is the arithmetical operations which we may carry out with confi- e as a prescription of long-term policy when we can legitimately me that Uspensky's two-fold condition does hold good suggest a n and ready pattern for short-term policy in a changing situation the firm fixes its stake with a sufficiently generous safety margin.' he second take the sentence 'This new theme which now invites scrutiny is a class of problems which to writers of the century cedent to the promulgation of his eternal law of population by n Malthus might well seem to be cognate to the hazards of the ler.' How much simpler it would have been if he had said 'Our theme is a class of problems which seventeenth-century writers d have thought cognate to gambling.' Malthus's law of population elevant, let alone the fact that he was a parson. Complication is mes necessary in order to be precise, but unfortunately Professor n is scarcely ever this. It is rare for him to give a formal definition rm, he is usually content to rely upon example; and when he does n misses the point. Thus on p. 381 we find the sentence 'We speak of such a schema as *asymptotically* adequate *a priori* since is a finite probability that the experiment will *not* fulfil our ex- tions, if r itself is finite.' This is meant as a definition of asymp-

totically adequate, but is useless because it completely omits the vit qualification that the probability mentioned tends to zero as r tends t infinity. This is surely the major point of the definition. Incidentall non-zero would be better than finite.

The next chapter is 'The place of Laplace and the Grand Climacter of the Classical Tradition'. The law of succession is discussed and th usual point is made about the relevance of the class of situations t which the probability refers. Hogben argues that Laplace's argumer is essentially a two-stage one, changing prior frequency probabiliti of the first stage into posterior ones in the second.

After the historical first part, the second part is devoted to the Ca culus of Error and the Calculus of Exploration. 'The Normal La comes into the Picture' in the first chapter and the central lim theorem is again discussed. We then turn to the method of least squar and the concept of point estimation and consider these from bot Gauss's and Legendre's viewpoints. The method of maximum likel hood is also discussed. The next two chapters serve to draw a vali distinction between, on the one hand, the Gaussian theory of lea squares used in estimation and, on the other, the theory of regressio which does not consider true values disturbed by error but only aims t provide a description of the phenomena under observation. It is fact analysis's turn for criticism in the final chapter of this section.

The third section is brief, containing only two chapters, on Statistic Mechanics and Mendelism respectively. The final section on 'Th Calculus of Judgments' is the most substantial and in it we come to th heart of what the author considers to be the crisis. With importa reservations, the first three divisions of statistics meet with the author approval, but he is critical of this final calculus. The first chapt distinguishes between statistical prudence and statistical inference. Th author gives no definition of these terms and I cannot, with confidenc explain the distinction between them. But prudence appears to be th sort of thing an engineer has in mind when doing quality control, an inference, at least in the significance-test situation, refers to an invest gation into the validity of a null hypothesis. The approach is agai partly historical, including, for example, a discussion of Yule an Greenwood's famous paper on the statistical evidence about the effe of inoculations. The chapter ends on a relevant and interestin trumpet note of challenge: 'The onus lies on the exponent of statistic theory to furnish irresistible reasons for adopting procedures whic have still to prove their worth against a background of three centuri of progress in scientific discovery accomplished without their aid'. I the next chapter a simple genetical situation in which we have to decid between two possibilities is discussed. The concept of an error of th second kind is introduced in conjunction with the idea of an alternati hypothesis, and the author's main criticism is foreshadowed in a me tion of the effect of sample size on the inference.

Throughout the next few chapters the author makes extensive use

thesis-observation grids (H.O.G.). The rows of the grid refer to ypotheses, the columns to the observations, and an entry gives the ability of the observation of the column under the hypothesis to n the row refers. It provides a clear description of a family of ability distributions. There are several simple examples of these rating different facets of the statistical argument. The third ter of this section introduces the ideas of Wrighton, for whom the or obviously has a very high regard. Wrighton defines an experi- as *holonomic* if each column of the H.O.G. has a single non-zero . This is the ideal experiment because it has the property that observation leads to a complete identification of the hypothesis. re holonomic experiments are possible no statistical theory is sary. On the other hand non-holonomic experiments (called *atory*) need statistical methods in their analysis. An extension of olonomic idea is the concept of *a priori adequate*. An experiment is *ri* adequate with respect to a given set of terminal statements and en level of assurance of the correctness of these statements if, *a* you know the experiment is capable of reaching some one of these nents with the required assurance. Thus if the experiment is one rning a normal mean μ where the standard deviation is known unity, then a random sample of size sixteen is *a priori* adequate respect to terminal statements which give the value of μ to the st integer with confidence coefficient of 95%. (The confidence als will be $\bar{x} \pm 2/\sqrt{(16)} = \bar{x} \pm \frac{1}{2}$, where $\bar{x}$ is the sample mean.) On ther hand, if the standard deviation is unknown, the experiment *a priori* adequate for the same purpose because the length of the dence interval (involving the sample standard deviation) is not ctable in advance.

t we are anticipating Professor Hogben. In the fourth chapter of ection several H.O.G.'s are considered with a view to 'exhibiting rocedure and interval estimation as variants of a common pattern soning; making explicit certain canons prerequisite to the complete tion and prescription of a rule of stochastic induction'. Four of canons are stated on p. 432, the fifth canon is discussed in the chapter (p. 449). Part of it says that 'no terminal statement may a higher level of precision than the rule in its entirety can se', and is the basis of the objections levelled against the confidence al for the mean when the standard deviation is unknown, des- d above. The terminal statements made when the sample standard ion is small will have a higher level of precision (i.e. smaller , than the rule gives on the average. The chapter ends with a criticism of the analysis of variance technique. There then an Epilogue in which three theses are defended. Briefly these (1) only frequency probabilities are meaningful; (2) a rule of ical induction cannot be defended by citing what it does in any ular case but only by considering its action in the set of situations ich it can be applied; (3) a fixed framework of repetitions must

be introduced in any stochastic argument. There are four appendixe to the book; The Central Limit Theorem; The Umpire Bonus Mode (i.e. Regression Theory); Determinant Notation; and finally, Signifi cance as interpreted by the School of R. A. Fisher. The last has los much of its value because in the interval between writing and publi cation, Fisher has published his own book on the foundations. Ha Hogben had it available at the time he wrote the appendix he woul have found his criticisms much easier to make. The three remainin appendixes provide mathematical demonstrations of certain result without giving valid proofs.

It is dangerous to summarise a five-hundred page book in a few line but the author's arguments seem to amount to saying that statistics ca be divided into four calculuses in which frequency probabilities are th only relevant ones. In the calculus of judgements he follows th Neyman-Pearson-Wald approach, introducing errors of the second kind for example, and rightly rejecting the Fisherian approach; but he i dissatisfied with the former approach because of the indeterminancy o the final *type* of answer before experimentation begins. The subdivisio seems worth while, but it is doubtful if the arguments need have bee so complicated and to have used such accumulations of historica material. The use of the Neyman argument appears sound and th exposition of the technique through the H.O.G.'s is good (this is th best part of the book). The restriction to frequency probabilities an the criticism of Neyman in connection with the confidence interval associated with the t-distribution appear unjustified. Let us conside these two points.

If we have a random sample of a size n from a normal distributio of unknown mean and variance, and if $\bar{x}$ and s^2 are respectively th sample mean and variance, then we can say with confidence $(1-\alpha$ that the true mean lies in the interval $\bar{x} \pm t_\alpha s/\sqrt{n}$, where t_α is ob tained in terms of α and n from Student's t-table. Wrighton an Hogben object to this argument, apparently because it is incapable o leading to statements of prescribed precision: precision, that is, in th sense of limiting μ to some interval of prescribed length. Neyman, o course, has recognised this defect and Stein has produced a two-sampl method of reaching the prescribed precision which is, however, sti unsatisfactory because it ignores part of the data. Is the Wrighto Hogben objection one which suffices to destroy almost all the moder edifice of statistical theory, or is it merely a criticism of our inability t solve certain problems within the edifice? They argue the former: bu surely it is only the latter. If, instead of taking a random sample o size n from our unknown normal distribution, we sample according t a more complicated rule, R, that is sequential in form, then it is no inconceivable that for some R it would be possible to make confidenc statements about the mean of prescribed length. (Stein's two-stag method does this unsatisfactorily.) At the moment we do not kno

to find R in this case: we can do it if the variance is known (by a -size sample) or when there are only two possible values for the butions (by a probability-ratio test due to Wald). They might mately criticise the *inadequacy* of modern statistical techniques but it air to criticise the *basis* of those techniques that exist. It is like say- We do not know how to store electrical energy effectively, therefore hole theory of electricity is at fault.' The statistician cannot con- rules R but it does not follow that statistics is at fault—unless it be shewn that R does not exist. Actually the construction of R is difficult because of the restriction to frequency probabilities, so let w turn to this aspect of Hogben's argument.

objects to probability statements about hypotheses, that is prob- y statements which incorporate the notion of degree of belief (such John Cockcroft's about Zeta: 'I am 90% certain that the neutrons produced by a thermonuclear reaction'), because they do not fit h the behaviouristic outlook. They are not objectively verifiable e sense that the statement 'the probability of a coin falling heads s. The argument ignores the work of Savage (which is not men- d by Hogben). Savage's results have been enough to change my on probability and since they are not at all well known, even to etical statisticians, and since Savage presents the reader[2] with much ematical technicality, it may be convenient to summarise them We shall also try to indicate some of their simpler applications.

starts from a situation in which the world is in a number of ble 'states' and the statistician has to make one among a number sible 'decisions'. An example is provided by one of the situations dered by Hogben (p. 345) where the states are two in number, e animal is genetically normal, (*b*) it is heterozygous; the decisions act as if (*a*) holds, or as if (*b*) holds. This is the simplest two-state, ecision problem. The reader will have no difficulty in inventing complicated ones. The argument now rests upon considering general rules should guide the statistician in making his decision, face of the uncertainty he has about which is the true state of the . Savage proposes certain rules and then says that a man is nal' if he obeys these rules in his decision-making. He intends that rules should be acceptable to all people, they are the axioms of bject. It may perhaps help if I give one of the axioms that sses part of what Savage calls the 'sure-thing' principle, to show neral nature of it, and to illustrate the fact that the axioms are acceptable. Suppose there are two states s_1, s_2, and two de- d_1, d_2. Suppose the statistician says: 'If s_1 were true I'd certainly '. (If the state of the world is known the determination of the ecision is presumably an easy matter.) Suppose also that he can f s_2 were true I'd also take d_1'; then a rational man should surely even if he is not informed about whether s_1 or s_2 is true. The ement that he will act in this way is the 'sure-thing principle'.

Savage introduces several axioms of this sort, and some which ar mathematically convenient in the same way that it is convenient t speak of an interval of real numbers, $0 \leqq x \leqq 1$, even if x can only b measured to a limited accuracy. Neither group of axioms seems t contain any that can be seriously objected to, though one, which de mands a complete ordering of all decision procedures, may be difficu for some people to accept, though even more difficult for them to reject

Having defined his 'rational' man, Savage spends a large part of hi book proving the following result: a rational man always acts as if h had a probability distribution over the states of the world and a utilit function describing the utility to him of each decision in each state; h chooses his decision by maximising the expected utility. Some readers wi probably be familiar with the utility idea in connection with the writing of von Neumann and Morgenstern. They shewed that, subject to mil axioms, such a utility function always exists. Savage extends this ide and shews that, in addition, a distribution of probability also exist 'Exists' in these two sentences means that all the decisions which ar taken could have been reached with their use. This seems to be a essentially behaviourist outlook and should appeal to Professor Hogbe

In the little example above, the result shows that there are positiv numbers p_1, p_2, adding to one, giving the probabilities of the two state s_1, s_2; and utilities u_{ij} describing the utility of decision d_i when s_j obtains The statistician will choose d_1 if the expected utility obtained by s doing, that is $u_{11}p_1 + u_{12}p_2$, exceeds that obtained by choosing d_2 $u_{21}p_1 + u_{22}p_2$. If he is faced with another situation with the same state s_1, s_2 but different decisions to be made then the p_i's will remain un altered but the utilities may well be changed. If in the original situa tion additional evidence becomes available (if, for example, he per forms an experiment), the utilities will remain unaltered but th probabilities will be changed into others according to Bayes's theorem namely the probability after the experiment, usually called the posterio probability, is proportional to the product of the likelihood of the resul of the experiment and the probability before the experiment, usuall called the prior probability.

Before Savage's argument became available writers, such as Si Harold Jeffreys, had introduced probabilities without a frequenc interpretation in a way which did not seem to many people to be ver compelling; and they had some difficulty in shewing that the probabili ties obeyed the same mathematical rules as frequency probabilitie Now anyone who is to disagree with them has got to refute at least on of Savage's axioms (or find a flaw in his mathematics). Savage's argu ment also provides a means of investigating a person's probabilitie He has only to be presented with a betting situation with variabl rewards. Thus in the above illustration suppose he is offered odds o C to 1 against s_1 with two decisions, to accept the bet, d_1, or make n bet, d_2. The utilities for a stake of one unit, are $u_{11} = C$, $u_{12} = -1$ $u_{21} = u_{22} = 0$. The man will bet provided

$$u_{11}p_1 + u_{12}p_2 > u_{21}p_1 + u_{22}p_2$$

rovided

$$Cp_1 > p_2$$

$$p_1/p_2 > 1/C.$$

rying C until the bet is accepted the ratio p_1/p_2, and hence p_1 and ay be determined. It should be noted that the probabilities that ge shews to exist for the rational man are not necessarily capable e frequency interpretation that Hogben thinks is so necessary. To asise this I shall call them 'Bayes' probabilities because, frequency t, they are the ones needed for insertion into Bayes's theorem. ge argues that they are 'personalistic', that is, they are a property e individual and not of society. I would dispute this myself, and with Jeffreys in saying that in scientific questions they are ob- e. They only differ between individuals because the individuals fferently informed; but with common knowledge we have common ian probabilities. We can ignore this side-issue in the present nt.

summarise then: it would seem that Savage's axioms are reason- nd that, granting them, any person must act at least *as if* he had yes' probability distribution and a utility function. This is not to at he must openly use these ideas, but that any rule he uses for on-making must be consistent with the possession of them. We herefore investigate our standard statistical procedures to see if fulfil this requirement. One of them that fails to do so is that uses an unbiased estimate, i.e. an estimate which will, on in- te repetition, give, on the average, the right answer. This is ising, since it is one of the oldest ideas in statistics, but not disastrous se it can often be replaced by the modern idea of invariance— h doubts have even been raised about the propriety of this pt. For example, in the usual least-squares theory, instead of g about best linear unbiased estimates, we can refer to best ant estimates. But most of the concepts, such as power, fit into ility-prior probability framework and post-Savage statistics looks much like the pre-Savage version. However, there are some points ference between the usual approach and the utility-'Bayes' prob- y method that deserve mention.

first concerns the use of confidence intervals. Suppose that an menter has obtained some results and from them has been able duce a confidence interval for the parameter that he is interested ecifically, suppose that he can say with a confidence coefficient % that the parameter θ lies in an interval (a_1, a_2). Now there is no concerning the meaning of this statement: if the experiment had epeated indefinitely and the interval calculated each time accord- he fixed rule by which (a_1, a_2) was obtained, then, on the average, terval would have contained the true parameter value 95% of the

time. Indeed the parameter value may change from experiment t experiment, or even the whole experiment alter, provided only that valid rule is used for calculating the interval. This meaning is a per fectly sensible one within the usual frequency probability approach Consider, next, not the mathematical meaning of this statement tha θ lies in (a_1, a_2), but instead the meaning it has in any application to th real world, particularly to a non-statistician who is trying to appreciat it. Usually this amounts to saying that the experimenter is fairly sur that θ lies between a_1 and a_2; if he wanted to be more certain about hi statement the coefficient could be increased to 99% or higher, th interval would widen but he would feel more sure of his statement Indeed, if, instead of using an interval (a_1, a_2), he had chosen a singl number a_3 such that the confidence that θ was less than a_3 is 95% then we may write in an obvious notation

$$\mathrm{cf}\{\theta < a_3\} = 0{\cdot}95$$

or generally

$$\mathrm{cf}\{\theta < \mathrm{a}(\alpha)\} = \alpha \qquad \ldots . (1$$

where $\mathrm{a}(\alpha)$ is some function of α (typically $\mathrm{a}(\alpha)$ increases with α), an α lies between 0 and 1. Now these statements are very close to probabi lity statements about θ. If there was a probability distribution (eithe in a frequency sense or in a 'degree of belief' sense) for θ, then for any between 0 and 1 we would be able to find a function $\mathrm{b}(\alpha)$ such that

$$\mathrm{p}\{\theta < \mathrm{b}(\alpha)\} = \alpha \qquad \ldots . (2$$

Consequently there is an extremely close connection between confidenc ideas and Bayesian ideas, expressed by the similarity between (1) an (2). They both express some form of assurance or conviction about th true value of the parameter and they both do it through probabilit statements. However their meanings, within the mathematical syste are quite different. One relies on the complicated idea of rando intervals covering the true values and the other on betting ideas. It therefore difficult to see which of (1) and (2) should be used in practical situation.

One argument is to say that the confidence statements are only to b used when the experimenter is 'completely ignorant about the para meter' (whatever that means) or equivalently when a 'Bayes' distribu tion does not exist. But this argument is defeated by Savage's proo it always exists for a rational man. Another possibility is to hope tha perhaps (1) and (2) are the same. For example, suppose the experime has been performed and we have our confidence statement (1). The we have also a 'Bayes' probability (2), by Savage's argument, obtaine as a posterior distribution from Bayes's theorem starting with a pri distribution before the experiment began. Could it happen that f some prior distribution (1) and (2) would be identical? If so this pri distribution would correspond to the position of ignorance demande by the user of confidence intervals. Unfortunately it can be shewn tha

nly very rarely indeed that (2) can be the same as (1) for some distribution. (The case of the normal mean is a conspicuous and t isolated exception.) Furthermore, confidence statements do not ine according to the rules of the probability calculus: a point Fisher seemed to misunderstand when using fiducial arguments. xample, if the experimenter performs a second experiment the ination of the separate confidence statements for the two separate iments into a single statement for the combined experiment is ng like as simple as it is with the Bayesian statements, which y combine by Bayes's theorem. Consequently (1) and (2) are letely different, in general, and yet they both seem to satisfy the ical requirement that some statement about θ should be made an experiment has been performed. Which should be used? Is rong and the other right? If so, which is right? Is one relevant e type of situation, and one to another? If so, how can these ions be distinguished? The answers to these questions do not to be known, and if Hogben had wanted to find a crisis in statistical y, this is where he can find an interesting little difficulty to magnify ne. It can be argued that confidence intervals are relevant to nce problems and Bayesian methods to decision problems. But the only purpose of inference is ultimately to enable a decision em to be solved (as Hogben would presumably agree, because statements of knowledge are not as important as uses of knowledge ermining action) the argument would appear to fall to the ground. ere is a second difference between the approach to statistics via dence intervals and that via 'Bayes' probabilities which is important plications of just the type that Professor Hogben has in mind. d by use of the Bayesian argument the construction of a rule (that noted earlier by R) to accomplish his object is simple in principle h, in particular problems, it may lead to functions which are not ated at present. To understand this we recall the close connection xists between confidence intervals and significance tests. If, for value θ' of the parameter, we have a significance test at level of the hypothesis that $\theta = \theta'$, then by a well-known inversion ment we can construct a confidence interval (a_1, a_2) for θ. The al (a_1, a_2) contains those values of θ' which the test of $\theta = \theta'$ would y as non-significant. With this intimate connection in mind let sider what happens in a significance test. We take some statistic, s the sample mean, or t, the ratio of it to its estimated standard ion, and consider the probability that the value of the statistic be greater than or equal to the particular value we happen to the experiment. In other words we not only consider the value at we have but also the other values of t that we might have had, ght have if we do the experiment again. Or, to put it more ily, we not only consider the sample we have but also the samples ght have had. For example when our experiment has yielded n $x_1, x_2, \ldots x_n$ under conditions which we believe make them

form a random sample from some partially unknown distribution, consider all the random samples of the same size that we might ha had from the same distribution. In order to calculate the probabil for our significance test we integrate over some of these samples, example, over those with a numerically larger value of t than obtained. Contrast this method of analysis with that based on pr probabilities which are Bayesian. The observations enter the calcu tions only through their insertion in Bayes's theorem, and here th only enter through the likelihood function. The likelihood functior in the discrete case the probability, and in the continuous case t probability density, of the observations actually obtained, consider as a function of the unknown parameters. Consequently it is only t probability of the observations actually obtained that matters to us, t observations that might have arisen do not concern us except insofar they determine the likelihood function. Thus a confidence approa (or a significance test argument) involves considering samples oth than those actually observed, and the Bayesian approach does not. example may clarify the issue.

Let n independent observations be available, each of which is eith 0 or 1, and the probability of being 1 is p, the same for all observatio Then the probability of the particular sequence that we have $p^r(1-p)^{n-r}$ if it happens that r of the observations are 1 and $n-r$, (We are writing down the probability of the sequence actually o tained, not the probability of obtaining the value r. Several sequenc ${}^nC_r$ in number, of which ours is one, give the same value of r. T usual ${}^nC_r$ is therefore omitted from the expression.) Next let us supp that this *same* sequence has been obtained independently by t scientists. One of them settled on a value of n and took n samp obtaining r 1's; the other settled on a value of r and took n samp before he reached r values that were 1. (The second scientist is adopti Haldane's inverse binomial method.) Now suppose each scientist p poses to make a significance test at level $1-\alpha$ of the hypothesis th $p=p_0$, say. They will proceed as follows.

Scientist 1. He will calculate the smallest value R such that if value of p is p_0 the probability that r will exceed R is, for fixed n, greater than $1-\alpha$. In order to do this he will evaluate

$$\Sigma p_0^r(1-p_0)^{n-r}$$

over all sequences, in all possible orders, with $r>R$ and fixed n. will say that the result is significant at level $1-\alpha$ if the observed va of r is greater than R. (I am assuming he is only interested in a possi *increase* of p above p_0, and I am ignoring the fact that the sum will be exactly $1-\alpha$ because of the discrete form of the distribution: neit point is relevant here.)

Scientist 2. He will calculate the largest value N such that if the va of p is p_0 the probability that n will be less than N is, for fixed r, greater than $1-\alpha$. In order to do this he will evaluate

$$\Sigma p_0^r(1-p_0)^{n-r}$$

ll sequences, in all possible orders, with $n<N$ and fixed r. He y that the result is significant at level $1-\alpha$ if the observed value less than N.

arly these two significance tests, and therefore the resulting confi- intervals, are entirely different. If the two scientists can now be ded to become Bayesians for a moment and agree on their prior ution with probability density $f(p)$, then their posterior distribu- ill be $Cf(p)p^r(1-p)^{n-r}$, where C is a constant, for both of them, he likelihood is the same for both. I leave the reader to think out nself the implications of these two viewpoints in this situation. usly the confidence approach has something to be said for it e it does take account of the different experimental approaches y the two scientists, whereas the Bayesian method does not. On ner hand here are n observations honestly obtained, what does it how they were obtained? By honestly, I mean that no sup- n of data has been allowed, both scientists reported honestly hey saw. The situation is clearly changed when any 'cheating' duced. Incidentally we can now see why the concept of unbiased cs does not find favour in the Bayesian approach. Scientist 1 will as his unbiased estimate of p, scientist 2 will use $(r-1)/(n-1)$ as ut if they were Bayesians they must agree.

fact that a Bayesian (if I may call him that) needs only the od to reach his distribution for θ, and does not need any sum- or integration over the sample space, makes the construction of ng schemes to reach desired types of terminal statement (to Hogben attaches such importance) an easy matter. The Bayesian vith his prior distribution and continues taking samples (one at a he can, or more if convenient) until the resulting posterior dis- on is to his liking, that is until he can make the terminal statement required type. The only two problems are (i) to calculate the utions and (ii) to say what feature of the distribution one is ted in. The first is purely an analytic problem. The second is One possible way to proceed is to continue until 95% of the or distribution is contained within an interval of preassigned The choice of 95% is ambiguous and it is perhaps better to use escriptive statistic of the distribution such as its variance, and ask at be sufficiently small. One possibility is to use Shannon's in- ion measure for the statistic and to continue until this is large This is equivalent to the approach based on the variance he sample size is large. It is quite easy to justify Haldane's binomial method (the method of Scientist 2 above) by this ure.

amount of reading that has been incorporated into this book is ut it is mostly reading of the older writers. The newer work, larly of Savage, has been ignored. This is perhaps the main

defect of a book which correctly assesses the need for a revaluation
statistical ideas, but does not quite succeed in placing the emphasis
the correct place.

REFERENCES

[1] Hogben, L. (1957). *Statistical Theory*. George Allen and Unwin Ltd., Londo
Pp. 510. 45s.
[2] Savage, L. J. (1954). *The Foundations of Statistics*. Wiley, New York; Chapm
and Hall, London.

NOTES AND COMMENTS

March issue of *Applied Statistics* included an account by Mr E. D. van f recent developments in quality control. It may be of interest to our s to learn that since that issue the UK, through the British Productivity il, has decided to become a full member, for at least a trial period, of ropean Organisation for Quality Control. Consideration is now being c means whereby British industry can obtain the maximum advantage h this country's participation in the activities of the European sation.

organisations or individuals who are interested in the work of EOQC vited to communicate with Mr V. M. Green, Secretary, BPC Quality ol Committee, 21 Tothill Street, London S.W.1.

* * *

interesting to note that during the last twelve months no less than five cal conferences have taken place, in addition to the usual activities of ciety and its Sections. In December 1957 the London Group of the rial Applications Section arranged a half-day conference on Pro- n Control, and in May 1958 the Birmingham Group held a one-day ence on Statistical Sampling in Industry. The Study Section organised -end conference in March 1958; this dealt with Business and Economic ics and took place in Cambridge. A conference on Mathematical ics in St. Andrews occupied a long weekend in August; it was sponsored Research and Industrial Applications Sections. Finally there was a nd conference in Manchester, organised jointly by the Manchester ical Society and the Royal Statistical Society, in September. For the portion of the time the meetings were arranged in parallel sessions g with Economic and Social Statistics and with Industrial Applications. se conferences, with their opportunities for formal and informal dis- ns, were greatly appreciated by all who took part. It is a healthy sign o many Fellows and Members of Sections are willing to devote their o these activities, and our especial thanks should be extended to those rganised them.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT GROUP

19th March 1958. Comparison of Industrial Efficiency: L. H. C. Tipp and J. C. Pritchard. Joint meeting with the Birmingham Productiv Association.

2nd April 1958. The Planning of Experiments: D. R. Cox

7th May 1958. One-day Conference on Statistical Sampling in Indust A Survey of Statistical Sampling for Management: E. D. van Re Statistical Sampling in the Control of Quality: R. H. S. Lesser. W Sampling: F. S. Campbell. Sampling in Accountancy: H. C. Mackenz

EDINBURGH GROUP

23rd April 1958. Errors in Choosing an Elite: B. Woolf

LONDON GROUP

10th April 1958. Sophisticated Chemistry: J. Youden

2nd May 1948. Component Replacement: J. Taylor

MERSEYSIDE GROUP

15th April 1958. Stochastic Linear Programming: S. Vajda

NORTH-EASTERN GROUP

12th March 1958. Some Applications of Statistics to Research on Anim Nutrition: A. W. Boyne

SHEFFIELD GROUP

24th April 1958. Monte Carlo Methods: K. D. Tocher

STUDY SECTION

LONDON GROUP

9th April 1958. Economic and Statistical Problems of Airline Operatio S. F. Wheatcroft

21st May 1958. Long-Term Demands for Fuel: E. F. Schumacher

MEDICAL SECTION

29th April 1958. Some Current Problems in Demography: P. R. Cox

27th May 1958. Mortality from Lung Cancer in Relation to Populat Density: R. E. Waller

BOOK REVIEWS

rtin: *Social Aspects of Prescribing*. London: William Heinemann Ltd., 1957. Pp. 30. 21s.

the past few years the increasing cost of medical prescribing in the al Health Service has been a matter of growing concern. Because an independent committee has been appointed under the chairman- Sir Henry Hinchliffe to investigate the factors contributing to the cost. This treatise forms the first evidence to be submitted to the ttee on what is a complex problem.

contents are arranged in two main parts, supported by a number of ices. The first part traces the historical background of prescribing 00 up to 1951, and also describes the administrative control exercised the National Insurance Scheme and, more recently, under the al Health Service. The second part covers the analysis of prescription s and the changes in prescribing between 1947 and 1951.

earliest attempt to limit the drug bill was the procedure instituted he National Health Insurance Act of 1912, whereby 'excessive' pre- was penalised. The methods then adopted have largely persisted present day. With the rapid growth of the pharmaceutical industry increased range and complexity of drugs, doctors have given up aditional role of dispensing medicine. It is therefore difficult for know exactly how much their prescriptions cost. But this is not factor which must be considered. The extent of medical services le in different areas and the growing importance of diseases charac- of old age must also be taken into account.

attempt to keep the costs within the pharmaceutical budget, the y of Health uses both educative and coercive methods which take different forms, and are applied to doctors and patients. There is rly defined boundary between the two approaches, and although ion sometimes ends in financial loss, this eventuality is rather remote. gh not expressly stated to be so, the Ministry's control is largely onal, and maximum stress seems to be placed on the information visory work of the Regional Medical Officer.

tempting to analyse the results of this policy no special survey was out. Instead, use was made of the mass of information already avail- out prescriptions and collected for administrative purposes, covering xecutive Council Area in England and extending over some twenty This procedure had several distinct advantages, perhaps the most nt of which was to demonstrate what use could be made of the ion by the administrator in this particular branch of the National Service.

seven medium-sized boroughs in England were selected and all ables that had the remotest connection with prescribing were ed and analysed. For convenience these were grouped under six adings: (1) prescriptions; (2) doctors' practices; (3) patients; (4) nditions; (5) geographical location; and (6) morbidity (ill health). ain group was subdivided, and altogether there were fifty-seven

such groups. The source of each variable is clearly stated but only tho based on unpublished data are included in the book. The statistical tec niques used were the product-moment correlation (r), the phi coefficie ($r\phi$), matrix of concomitance, factor analysis, multiple regression equatio and quartiles.

This analysis showed, among other things, that the frequency of pr scribing depended on the region, associated with climatic conditions, mo bidity, and local 'custom'. The local 'custom' was so strong a factor tha in the author's words, 'in spite of changes arising from the introduction the National Health Service, one could probably get as good a predicti of the average frequency of prescription per patient in 1951 by using t comparable figure for 1933 as by using any amount of contemporary info mation'. Average cost per prescription is closely related to social conditio as are the average sizes of doctors' practices and the extent of morbidity.

The book itself is well arranged, lucidly written and easily read. It forms most useful piece of social research, of interest not only to the social scient and administrator, but also to the sociologist and statistician. A. BA

Dudley J. Cowden: *Statistical Methods in Quality Control*. Englewood Cliffs, New Jerse Prentice-Hall; London: Bailey Bros. and Swinfen, 1957. Pp. xxiv × 727. $12.

Those who claim that there are scarcely any of the many statistical metho which have not some application to the control of manufacturing process will be encouraged to find a book of this scope with this title. The conte list, in which every paragraph heading is listed, occupies sixteen pages. On in a nation of prodigal publishers could this happen. Its very prodigality ha however, its disadvantages; for though the book may contain all that 'quality control practitioner', or so-called 'analyst', may require, it is dou ful whether there are many such equipped to master this book. Althou mathematical proofs of major theorems are relegated to appendices, t main text is liberally occupied with derivations of formulae (in which an u fortunate number of small errors occur); nor is the reader spared the mo recondite parts of the subject such as k-statistics and Gram-Charlier curv It is, in fact, the text of a formal course in statistical methods directed at stages towards methods for control of quality.

In this country, and one would have thought elsewhere as well, the cont of quality is either the responsibility of an engineer with only an elementa knowledge of statistics or else there is in the department a qualified stat tician who can cope with the out-of-routine problems that arise. This bo is too difficult for the first; he is likely to be defeated by the very number the forty chapters. However, the need is admittedly different in differe countries and if one were training statisticians solely as quality cont analysts there is something to be said for a special book.

The early chapters follow a conventional development of the subject statistical methods. Some of the concepts which are so necessary even at early stage are not developed as clearly as they might be, and are in mo elementary books. The confusion that is still prevalent about probability mentioned and the subject left at that. In my opinion the usual probabil statement involving confidence limits needs explanation and the stud profits from an expanded explanation. A straightforward explanation o significance test can be given without introducing the concept of errors the second kind.

s more especially after Chapter 10 on the analysis of variance that the
er headings take on the flavour of control, and the idea of process
ol is used to introduce likelihood-ratio, tests of homogeneity of variance,
ution of range, operating characteristics, systems of frequency curves,
egression. There is a useful chapter on the economics of control charts
ne rest of the book is devoted to sampling plans whose thorough treat-
includes the hypergeometric case, multiple-sample, sequential, and
uous-product sampling schemes.

important task for the quality control analyst is the design, conduct,
nalysis of experiments aimed at removing faults in production processes,
ere is no mention of the highly developed subject of experiment design.
se this is so bound up with significance tests and the analysis of variance
is no reason for regarding the subject as one to be studied separately.

E. D. VAN REST

n Feller: *An Introduction to Probability Theory and its Applications, Volume I* (2nd Edition). York: John Wiley and Sons; London: Chapman and Hall, 1957. Pp. xv × 461. 86s.

e first edition of this book, which appeared in 1950, quickly became
ished as the standard reference and text book in probability theory,
as since played a key role in stimulating research in this subject.

e new edition follows the same general pattern as the old, and, at first
, appears to be not much altered. In fact, it has throughout been fairly
ively re-written and streamlined, and the type has been re-set. This
abled Professor Feller to pack additional material into a space which
s of the first edition will recall was fairly tightly crammed with ideas.

ere are two new chapters, in each of which a topic previously treated
t of a chapter is specially singled out for extended treatment. The first
les an interesting new account of fluctuation phenomena in coin-tossing.
heory is now related to the classical ballot problems of combinatorial
bility, and a link with recent work on the fluctuation theory of sums
ependent random variables is indicated. Mathematical Induction
es the method of generating functions. This has made it possible to
he chapter near the beginning of the book, and by use of quite simple
ds, to shock the reader early on with astounding results at variance
ommon sense. The arc sine laws relating to long leads are illustrated
analysis of the results of a sampling experiment in coin-tossing. The
(short) new chapter now provides a unified account of compound
utions and branching processes. It includes mention of infinitely
le distributions (in the discrete case), and the relevant basic theorem.

most important changes in the book relate to an increased emphasis
unity of recurrent events and Markov chains, the latter theory pre-
g a new feature only in that it deals simultaneously with a set of re-
t events. In line with this, the terminology has been slightly changed:
ere previously called *certain* or *uncertain* events are now termed *persistent*
ient. This change permits the labelling of the states of a Markov chain
as persistent or transient, and avoids a previous slight inelegance.
tion is obtained by introducing the concept of a *waiting time* at the
ing of the book. A waiting time is the number of trials up to the
nce of a given event. A waiting time for a recurrent event is the same
recurrence time. (The concept of the same name used in queueing
should be distinguished; the author does not deal with this.) It is

pointed out that since the appearance of the first edition the basic theorem in Feller's theory of recurrent events has been discovered to be equivalent to an ergodic theorem of Kolmogorov.

The restriction of the treatment to discrete random variables is maintained in this edition. The justification seems to be that it permits a high standard of rigour with only elementary mathematical techniques. The optimum compromise here is, as always, a matter of individual judgement: certainly the artificiality of the restriction becomes apparent whenever the author discusses events connected with unlimited sequences of trials, the strong law of large numbers, and so on. He is also hampered in his treatment of stochastic processes. The normal approximation to the binomial distribution is obtained by means of Stirling's formula, characteristic functions not being available. (In similar circumstances I have used, for illustrative purposes in teaching, the approximation of the binomial difference equation to the normal differential equation.)

The book closes with a brief discussion of the theory of the Kolmogorov differential equations for stochastic processes. An explanation for the non-uniqueness of the solution in certain circumstances is given, based on recent work, and the link with diffusion theory is indicated.

The only theoretical error which I noticed in the previous edition has been put right (p. 407). A number of misprints have been corrected. The expected number of returns to the origin in the three-dimensional random walk is again quoted as 0·53. It seems worth pointing out that the value was recomputed by Domb (*Proc. Cambridge Phil. Soc.* **50**, 586) in 1954 and found to be 0·516 39 to five places.

Since the first edition several other books have appeared dealing with aspects of probability theory, including particular applications such as stochastic epidemics and queueing theory. There is still, however, no rival to Feller's book as a comprehensive introduction, at a reasonable level of abstraction, to the modern theory of mathematical probability and its applications: it may certainly be read with profit by all students of statistics. The price is almost twice that of the previous edition.

F. G. FOSTER

PUBLICATIONS RECEIVED

Lancelot Hogben: *Statistical Theory—The Relationship of Probability, Credibility and Error*. London: George Allen and Unwin Ltd., 1957. Pp. 510. 45s. (Reviewed in the article commencing on p. 186.)

A. B. Neale and Jean D. Wheeler (Editors): *City of Birmingham Abstract of Statistics, Number Four 1955-1957*. Birmingham Corporation, 1957. Pp. xii × 147. 10s. 6d.

This well-produced volume maintains the high standard set by its predecessors. It is hoped that the publication will now appear regularly every two years. Opportunity has been taken in this issue of printing some otherwise unpublished information gained in the 1951 Census of Population, and there has been some rearrangement of the chapters. Where possible, figures are given for 1957, but many go only up to 1956.

Mary Jean Bowman (Editor): *Expectations, Uncertainty, and Business Behavior*. New York: Social Science Research Council, 1958. Pp. vii × 202. $2.

INDEX

ıce sampling, a guide to, 133
. J., 72
ology, discrimination in physical, 63
, P., 1
sale, statistics of, 42

I. T. J., 104
, 202
. G., 70
66, 67
V. A., 128
Lee, C. M., 64
otting tests, variability of, 1
Computer Society, 126
Economic Statistics, An Introduction
review, 76
Productivity Council, 60, 199
t, S., 86
am, R. A., 70
. E., 129

accidents, statistics, 128
ous processes, control of, 63
gression control chart for, 48
D. J., 202
R., 129

E., 77
D. C., 42

c trends, official statistics of, 15
st, his statistics, 76
J. H., 62
forecasting, 129
mental hazards, interpretation of, 62
ological approach, 63
n Organisation for Quality Control,
199
V. D., 63
ents, planning of, 129

alysis, 114
tion of Numbers, The', review, 70
L., 203
G., 204
ons of statistics, 186
K., 25

V. A., 173
G., 58
r, U., 68

Hamaker, H. C., 149
Healy, M. J. R., 63
Higgins, I. T. T., 63
Hogben, L., 186

Ingram, G. I. C., 1
International Statistical Institute, conference, 126

Jessop, W. N., 132
Jowett, G. H., 114

Kemp, K. W., 63
Knowelden, J., 64
Kramer, C. Y., 173

Li, J. C. R., 71
Liddell, F. D. K., 128
Lindley, D. V., 186
lognormal distribution in quality control, 160
Lyle, P., 132

Mansfield, E., 48
marginal cost, 96
Martin, J. P., 201
Maurice, Rita J., 15
McCarthy, P. J., 130
medical treatment, assessment of, 64
Milner, E. S., 128
missing observations, analysis of variance with, 173
Morris, W. T., 65
Morrison, S. J., 160
mortality rates, 86
motivation research, 129
Muir, A., 96

Neyman, J., 67
'Numerical Methods', review, 70
Nuttall, C., 130

optimisation problems, computing methods, 64

Pickard, D. H., 129
Pounce, R. J., 129
'Probability Theory and its Applications, An Introduction to', review, 203
probits, application to observational data, 25
'Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol V', review, 67

'Rationalization of Industrial Decision Processes, The', review, 65
'Reduction and Presentation of Experimental Results, The', review, 66
'Regression Analysis of Production Costs and Factory Operations', review, 132
regression control chart for costs, 48
Reichmann, W. J., 70
Richardson, J. T., 66
Rosenblatt, M., 68

sampling inspection by attributes, basic principles, 149
Shankleman, E., 76
Smith, F. T. M., 71
'Social Aspects of Prescribing', review, 201
Stanley, P. J., 131
'Statistical Analysis of Stationary Time Series', review, 68
'Statistical Inference, Introduction to', review, 71
'Statistical Methods in Quality Control', review, 202
'Statistical Reasoning, Introduction review, 130
'Statistical Sources for Market Resear review, 130
'Statistical Theory—The Relationship Probability, Credibility and Er review, 186
statistics, foundations of, 186
stocktaking, statistical approach, 128
store-size, minimisation of, 104
surnames, initial letters of, 58

television, effects of, 128
Tippett, L. H. C., 132, 133

unemployment statistics, 129

van Rest, E. D., 60, 68, 203

Walker, A. M., 70
Wein, H. H., 48

Printed in Great Britain

Nos. 1 and 2 at
THE KYNOCH PRESS
Birmingham

No. 3 by
ROBERT CUNNINGHAM AND SONS LTD.
Alva

The £.s.d. of statistics

Increased and increasing demand for the Monroe-matic 8N1-213 calculator enables us to announce a substantial price reduction—from £498 to £410. Specially designed for statisticians and scientists and embodying numerous features of great value in their work, the Monroe 8N1-213 is now competitive in price with many far less advanced calculators.

For further particulars or to arrange a demonstration, please write to your local Monroe branch or to our Head Office.

MONROE

Monroe Calculating Machine Company Limited
Bush House, Aldwych, London, W.C.2
COVent Garden 0211
Branches in Birmingham, Bristol, Cardiff, Glasgow, Leeds and Manchester.

GRIFFIN

THE FUNDAMENTALS OF STATISTICAL REASONING

M. H. Quenouille, M.A. Research Techniques, L.S.E., London.

This book sets forth the underlying ideas and concepts which form the basis of statistics: in the author's own words, it deals with "the principles rather than the proofs or the practices" of the science, and will interest both the mathematician and the student of applied statistics
"Griffin's Statistical Monographs. . . ." No 3. 24s. net

THE MATHEMATICAL THEORY OF EPIDEMICS

N. T. J. Bailey, M.A. Reader in Biometry, University of Oxford.

General introduction—Historical sketch of mathematical epidemiology—Epidemiological principles—Deterministic theory—Stochastic theory: continuous-infection models — Stochastic theory: chain-binomial models—The measurement of latent and infectious periods—Recurrent epidemics and endemicity—The detection of infectiousness—Summary and conclusions—Index
36s. net

CHARLES GRIFFIN & COMPANY LIMITED
42 DRURY LANE, W.C.2
LONDON

BINDING

We shall be glad to undertake the binding of separate parts of *Applied Statistics* into complete volumes at a cost of 20s. 3d. per volume.

Binding cases can be supplied separately at 5s. 6d. each.

Please send enquiries to Binding Department.

OLIVER AND BOYD
Tweeddale Court, Edinburgh, 1

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part
20s.

Annual Subscripti
Post Free £4 4s.

Principal Contents of Volume 121 Part 3 1958

Industrial Production and Steel Consumption. By Sir Robe Shone and H. R. Fisher. (With Discussion)

Some Problems in the Measurement of Price Changes wi Special Reference to the Cost of Living. A Discussi opened by Dr. S. J. Prais.

The Design of a Long-Term Sampling Programme to Measu the Hazard associated with an Industrial Environment. J. R. Ashford.

Reviews of Books, Statistical and Current Notes, Additions Library.

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1
Telephone: Welbeck 7638